DEVELOPMENT AND APPLICATION OF NOVEL COMPUTATIONAL INTELLIGENCE TECHNIQUES TO THE MULTIVARIATE ANALYSIS OF METABOLOMICS BIOFLUIDS DATASETS

Mamadou Bamba

Computer Science and Pharmacy

De Montfort University Leicester UK

A Doctorate Thesis Submitted to

De Montfort University
In partial fulfilment of the requirements for the
Degree of Doctorate of Philosophy

December 2017

ACKNOWLEDGEMENT

I will never think a lot the Almighty ALLAH (GOD) for giving me the strength, the courage, the determination and the drive to complete this PhD thesis.

I would like to think then after my supervisors Prof. Martin Grootveld, Prof. David Elizondo and Dr. Randolph Arroo for their constant support and the time dedicated to the current thesis. Without their constant cooperation, their help and above all their guidance this work would have never reach a successful conclusion and I am really grateful.

I would like to give a very warm thank my family for constantly believing in me and giving the necessary support to me and my children, especially my sisters and brothers in the Ivory Coast, in France and in America. I know that without their encouragement I would have at some point give up this fantastic and tremendous experience that constitutes the PhD Journey.

I would like to thank my children for being patient with and to forgive for not being present for their education all the time during this journey. This PhD is a special gift to Bamba Idriss Amine and Bamba Adja Fatim, and I pray the ALLAH the almighty makes them stronger and brighter than their parents.

Finally, I would like to dedicate this thesis to my late parents who have who have given me the correct education and time during their lifetime to be strong, honest, humble be respectful to my teachers and always strive to complete a task assigned to me. I pray for them that ALLAH (GOD) give them rest in peace.

"Give me a place to stand and I will move the worlds"

Archimedes

ABSTRACT

The present decades have witnessed major advances in the development and applications of Computational Intelligence Techniques (CITs), which are commonly associated with metabolomics and omics analyses related to diseases diagnosis. This includes, amongst others, research work performed on Niemann-Pick class 1 and 2 diseases (NPC1 and NPC2 respectively), the severest form of which may involve liver dysfunction. Some of the main reasons for the high frequency of CITs use in metabolomics studies are also related to the development of techniques to detect major discriminatory metabolite variables for the purpose of disease diagnosis and progression. Alongside this, is the major demanding requirement to further understand potential metabolic pathways involved in order to improve our understanding of the molecular mechanisms underlying NPC1 disease.

NPC1 is a rare neurodegenerative disorder attributable to NPC1 gene function loss, which causes adverse fat storage at the lysosomal levels (Mathieson, 2013; Xu et al., 2010). However, plasma metabolite profiling can provide insights into disease diagnosis and prognosis, while providing a clear 'picture' of the underlying metabolites altered during disease processes, including their early stages. Currently, biomarker discovery appears as the most effective solution to employ regarding the monitoring of disease progression (Mathieson, 2013; Ruiz-Rodado et al., 2014).

In the present thesis, the intelligent tri-modelling techniques (ITMTs) which are combination of CITs applied to the multivariate (MV) analysis of biofluid datasets is proposed. The ITMTs serves as a combination of the scalar visualisation algorithm (SVA) for data visualisation and high-dimensional data representation into bi or tri-dimensional spaces. The optimum super support vector machine (OSVM) is also employed for the MV classification of metabolic datasets. Moreover, principal component regression (PCR) was also employed for data

probabilistic classification and regression purposes. This was followed by investigations of correlations between these biomolecular diseases features. Furthermore, the tri-ranking techniques (TRTs) was developed in order to establish a ranking between the NPC1 disease features, in addition to those available for the NPC1 liver dysfunction disease features in a mouse model system to determine their importance in the further development of these diseases.

High-resolution proton nuclear magnetic resonance (^{1}H NMR) is used as high-throughput multicomponent analytical technique to generate very large quantities of metabolic data, which hold essential and useful information regarding the metabolites analysed. Prior to the performance of MV metabolomics analysis, a robust data handling technique based on balancing the dataset, feature selection, and stratified cross-validation of datasets is involved. Furthermore, the intelligent task technology fit theory has been proposed here, enabling a swift, consistent and rational model development through threshold settings for model validations.

Application of the intelligent tri-modelling techniques (ITMTs) using SVA, OSVM and PCR, combined with the tri-ranking techniques (TRTs) have allowed the discovery of major discriminatory variables for NPC1 disease. Hence, using the blood plasma dataset the scalar visualisation algorithm could diagnose NPC1 disease through the following potential biomarkers: hexacosanoate acid, (R)-3-hydroxybutyrate, L-fucose, lactate, 3-hydroxyisovalerate, Citrate, N-acetyl-4-O-acetylneuraminate, methionine, and glutamine. Additionally, the SVA strategy highlighted the following major biomarkers in the ^{1}H NMR NPC liver dysfunction dataset, including glycogen, glutamate, glutamine, taurine, glycerophosphocholine, acetoacetate, taurine, myo-inositol, lactate, leucine, isoleucine, and alanine.

However, the PCR approach established a significant correlation between biomarker features for NPC1 disease, in addition to the mouse model of NPC liver dysfunction progression. Moreover, the OSVM technique could clearly segregate between the two classes of patients/animals in both disease pathogenesis studies.

This thesis presents the Intelligent Tri-Model Techniques (ITMTs), using ^{1}H NMR-linked metabolite profiling, new biomarkers for NPC1 disease diagnosis, and the NPC1-based liver dysfunction were discovered; these biomarkers displayed very high-performance accuracies.

This may represent a major advance regarding the diagnosis of NPC1 disease and its pathological sequelae. Such biomarkers may serve as valuable assets for monitoring the effectiveness of appropriate treatments for this debilitating condition, for example miglustat.

KEY WORDS

NPC1 Disease; NPC liver dysfunction Disease (NPC LDD); Biomarkers; Multivariate Analysis; Pathways Analysis, Visualisation, Scalar Visualisation Algorithm, Classification, support Vector Machine, Regression, Principal Components Regression.

USED ABBREVIATIONS

1D NMR	1 Dimension Nuclear Magnetic Resonance
AAM / 2AM	Algorithm for Alignment Model
AFM	Alignment Function Model
CHD	Coronary Heart Disease
CITs	Computational Intelligence Techniques
CLF	Chronic Liver Failure
CSF	Cerebrospinal Fluid
CUDA	Compute Unified Device Architecture
ECG	Electrocardiogram
FBI	Fit-Be-Intelligent
FDA	Food and Drug Administration
FID	Free Induction Decay
FRS	Framingham Risk Score
GC	Gas Chromatography
GRS	Genetics Risk Scores
GSI	Gonadosomatic Index
HBV	Hepatitis B Virus
HDL	High Density Lipoprotein
HP- β –CD	2-hydroxypropyl- β-cyclodextrin
IAD	Intelligent Algorithm Development
IITFM	Intelligent Technology Task Fit Model
IMP	Intelligent Modelling Process
ITMTs	Intelligent Tri-Modelling Techniques
LC	Liquid Chromatography
LR	Logistic Regression
MC	Monohexosylceramides
MGS	Miglustat
MLR	Multiple Logistics Regression
MRI	Magnetic Resonance Imaging
MS	Mass Spectrometry
MV	Multivariate

NMR	Nuclear Magnetic Resonance
NPC1	Niemann-Pick Class 1
NPC LDD	Niemann-Pick Class Liver Dysfunction Disease
OSVM	Optimum Support Vector Machine
PCA	Principal Components Analysis
PCR	Principal Components Regression
PCs	Principal Components
PLS DA	Partial Least Square - Discriminant Analysis
QTc	Q waves – T waves
SELDI-TOF	Surface Enhanced Laser Desorption Ionization – time of flight
SVA	Scalar Visualisation Algorithm
TRTs	Tri-Ranking Techniques

TABLE OF CONTENT

LIST OF ALGORITHMS

LIST OF FIGURES

LIST OF FUNCTIONS

LIST OF TABLES

"Give me a place to stand, and I can move the world"

Archimedes

1. INTRODUCTION

1.1. Introduction

Research in general has always progressed incrementally, especially when new researchers generate new methods and techniques as improvements over existing ones, and or developing new ways of solving existing issues. This incremental development has seen tremendous changes occurring in different areas where researchers have been very active. This includes those such as mathematics, physics, robotics, computer sciences, medicine, pharmacy, etc.

The last three areas noted above are the ones especially featured in this research investigation, with many applications of metabolomics analysis techniques. These research fields include cancer, type 1 and 2 diabetes, Alzheimer's and Parkinson's diseases, NPC and further lysosomal storage diseases, etc., (Beger et al., 2010; Yun et al., 2016; Daniel and Forbes, 2015). However, the differing sets of biomarkers discovered have not really met the level of expectations sought. Indeed, such research findings have fallen short to provide ultimate diagnostic and prognostic biomarkers or to suggest appropriate treatments via subsequent drug targeting strategies. To date, no particular effective, disease-specific treatment is available for NPC1 disease (National Institute of Neurological Disorder and Stroke, 2016; Niemann-Pick UK, 2012).

This research work proposes CITs based on the intelligent tri-modelling techniques (ITMTs) approach in biomarkers discovery. The reason of such development is to employ three complementary techniques for dataset dimensional reduction, data visualisation, data classification, and finally a regression by establishing a correlation between disease features and the diagnosis of NPC1 disease.

The introductory component of this thesis encompasses the research background providing essential information on this metabolomics research, followed by the research motivations stating the reasons for conducting this study, and the research objectives highlighting the main questions supporting this investigation. Further research aims and objectives highlighting the main achievable tasks performed are given, followed by the contribution of this work to research, which presents the research achievements and successes. Finally, the different stages of this introductory chapter are detailed for clarity purposes.

1.2. Research Background

For a very long time in the history of humanity, research performed on human diseases have been a very daunting task for researchers, especially in the disease diagnosis fields of study where more support in terms of funding is needed. Research and development (R&D) for top neglected diseases was $2.956 billion in 2008, and $3.189 billion in 2009; while funding on the same diseases were $1.873 billion in 2008 and $2.121 billion in 2009. Further investment, such as total public funding (high income countries and multilaterals) on disease was $1.78 billion in 2008, and $2.049 billion in 2009; showing a drop in investment in the R&D 2008/2009 (Ngo et al., 2013).

In addition, it should be mentioned that diseases control, cure and eradication can be very difficult objectives to achieve, without sustained investments in the research areas. This meant that more effort needed to be done in this regard. Thus, investment in research related to neglected diseases, has constantly been above the US $3 billion from 2009 to 2014; while in 2014 there was an increase of US $150 million to reach US $3.37 billion due to the Ebola epidemic outbreak in the west African region (Sanderson, 2015).

Furthermore, researches showed that in some specific cases diseases genes have the potential of changing their structure through alteration, mutation, etc., which can be very challenging in terms of finding adequate treatments and appropriate remedies. This is valid for diseases like Niemann-Pick disease, which is rare progressive genetic disorder, with different forms, such as the Niemann-Pick type A, B, and C (National Institute of Neurological Disorder and Stroke, 2016; Niemann-Pick Disease Overview, 2013; Niemann-Pick UK, 2012).

Type C Niemann-Pick disease is an autosomal recessive inherited condition that is both rare and fatal. The Niemann-Pick class C (NPC) disease is present in the following ratios; class 1 or NPC1 (95%), and class 2 or NPC2 (5%) of the total infected population, i.e., with an incidence of approximately 500 cases diagnosed worldwide. The figure may be higher due to difficulties in diagnosing this medical condition (Blom, 2003; Maue et al., 2012; "Niemann-Pick Disease Overview," 2013). Research findings established that NPC1 patients carry the mutated NPC1 genes that encodes the protein NPC1, whilst their parents, also known as heterozygotes or carriers, have one functional NPC1 gene and one non-functional one. For NPC2 disease, this case is very rare, and shows no clearly observable signs or symptoms (Akst, 2014; Balch et al., 2008).

NPC1 disease can affect individuals before birth, but may not be apparent until late childhood. In fact, there is no clear age when certain manifestations become visible, including the first appearance of the symptoms, the disease's progression and its neurological manifestations (Maue et al., 2012; Yang, 2005). In addition, there are limited options concerning the NPC1 disease treatments and this lack of appropriate therapeutic method at early stages can lead to a more severe form of the disease (Fan et al., 2013). In fact, in view of the disease's link to gene mutation, NPC1 gene mutation screening has been applied, but failed to provide the necessary information regarding disease diagnosis and progression (Maue et al., 2012; Yang, 2005). Therefore, there is an urgent requirement for the proposal of appropriate treatment methods for this debilitating condition.

Consequently, the usage of computational intelligence techniques (CITs) in metabolomics multivariate analysis, for NPC1 patients' metabolites scrutiny appeared to have the potential to handle this challenge, especially through the detection of new biomarkers or the improvement of existing ones (Duarte et al., 2014; Ruiz-Rodado et al., 2014; Turkoglu et al., 2016 ; Wang et al., 2013). The need of using the computational intelligence techniques to produce efficient models in biomarkers discovery, especially in relation to finding effective and efficient treatments for the NPC 1 disease and the related liver dysfunction and associated disease is real.

Subsequently, the overview of the research background would ultimately lead to several motivating factors for the current study. These motivating elements supporting and sustaining this thesis are outlined below.

1.3. Research Motivation

Living organisms produce different type of biofluids such as saliva, blood plasma, urine, synovial fluids, etc., for several reasons including functionalities. In this case, the production is said to be functionally driven (MedicineNet, 2016; Novak et al., 2010; Rettner et al., 2016). Thus, most or all the secretions released by glands and cells are function specific substances. For example, urine is considered as the body system cleaner from all its impurities, whilst blood has different functions including carrying nutrients, and gases such as oxygen (O_2), and dioxide carbon (CO_2), protection, and regulation (MedicineNet, 2016; Novak et al., 2010; Oxford Dictionaries, 2016; Rettner et al., 2016).

Biofluids can act as indicator of the physical, mental, and psychical ability of a living being to adapt to its environment, because metabolites generated can be used to measure the degree of fitness, state of illness, infection level, etc., of the individual (s) under study, and its (their) suitability to live as normal being. Hence, how to tackle the disease diagnosis, progression, and improve the individual (s)' living condition could be investigated through biofluids metabolomics analysis.

Indeed, previous studies focused on diseases classifications using more conventional methods such as LC/MS, GC/MS, etc., combined with data analysis techniques such as principal components analysis (PCA), principal component regression (PCR), partial least squared discriminant analysis (PLS-DA), etc., have so far shown their limitations in terms of their efficacies in finding relevant biomarkers for diseases diagnosis (Amathieu, 2016; Becker et al., 2012).

The use of high-throughput techniques, including ^{1}H NMR spectroscopy has made it easy for researchers to generate large amount of data that carry countless information related to disease diagnosis, progression. In this respect, the scrutiny of this large amount of data through metabolomic multivariate analysis of biofluids, using CITs is an opportunity to grab the maximum knowledge on living organisms, and to examine their properties at molecular and or physiological levels. However, metabolite profiling would allow to capture information related

to the disease diagnosis and progression from an early stage, while designing ways for new biomarkers detecting and or improving existing ones.

This presentation of the different motivations sustaining the present study would be incomplete if some fundamental questions were not raised, giving clear orientation for future investigations to take place.

1.4. Research Questions

The research questions allow the research to be focussed, have a direction, set boundaries and they act as a frame of reference for assessing the work undertaken (O'Leary, 2013). The research questions help to achieve the research aims and objectives, while creating ways and platforms for arguments presentation (Duke, 2014). The main research questions that arose from this study are outlined below:

1. Can the use of CITs help in improving biomarkers discovery in the NPC1 disease and the severe form of the disease diagnosis and progression?

2. Can the use of CITs be an answer to diseases early diagnosis through metabolomics multivariate analysis of biofluids datasets?

3. Can the use of CITs help in establishing correlation between the NPC1 disease and the disease features on the one hand, and the NPC liver dysfunction disease (NPC LDD) and its features on the other hand?

4. Can the use of CITs help in establishing the interaction between a set of features and the NPC1 diseases at different stages of progression?

5. Can the biomarkers detection help in the understanding of the pathway followed and the diseases progression?

The present research questions ultimately trigger the following research aims that are hence exposed below.

1.5. Research Aims

Artificial Intelligence as one of the fastest-growing field in research, especially in the area of computer sciences has had many applications in various scientific and technological fields such as computational intelligence, bioinformatics, pharmacy, chemistry, engineering, etc. The application of CITs in the present study, has several aims which are outlined below.

Going through the different research questions highlighted above, it comes clearly that the aforementioned issues have not been properly addressed or not been addressed to a higher standard level. Indeed, in several studies conducted researchers while working toward biomarkers discovery, have not found the biomarker (s) that would have revolutionised the research area. Mostly, the biomarkers discovered would need further research to be conducted, in order to confirm the findings (Qi et al., 2012; Qiu et al., 2008). Nevertheless, the author of this thesis believes that this initial stage is as important as the final stage of definite validation of the biomarker discovered, because this is where the importance of the main discriminants is first established allowing further work to be carried out.

Therefore, although biomarker discovery offers much promise in the field of disease diagnosis, it has not yet attained and met the level of expectation the research community has placed on it; hence, it has not yet been employed to its full potential (Qi et al., 2012; Rocha et al., 2011; Wang et al., 2012). The following algorithms could be used as intelligent tools to address the issue, since their combination will enable the researcher to view the internal data structure, segregate between healthy and disease subjects, and finally, correlate disease features and disease diagnosis and progression.

The first among these algorithms is the scalar visualisation algorithm (SVA), developed as tool that converts scalar value to image. The ^{1}H NMR-detectable metabolite levels are transformed/assigned to index values. Hence, using a colour lookup table, a given index value is attributed and mapped to a colour in the colour lookup table (Johnson, 2015; Johnsona, 2012; Komura, 2016). In this manner, researchers can gain insight into more complex phenomenon and model them, including, for example, the underlying transformations occurring at the

molecular level which can, in turn, be visualised and understood (Johnsona, 2012; Kuhn and Johnson, 2013).

The second algorithm, the support vector machine (SVM), is a supervised learning algorithm used either for regression or classification purposes; but usually associated with classification problems. Thus, data points are plotted in an n-dimension space where n correspond to the number of features, with a hyperplane used for discriminating between diseased individuals and non-diseased ones (Ray, 2015).

Finally, principal component regression (PCR) is a statistical strategy that analyses datasets containing linear combinations of independent variables (principal components), which in turn generate outcomes in the form of binary or dichotomous variables. The possible outcomes might include; true or false, diseased or non-diseased, yes or no, 1 or 0 etc. The main reasons behind the model development is to establish the level of association of the independent variables with the outcome using the coefficient of regression r, or alternatively determine the strength of the relationship between two variables using α^2 the square of the coefficient of correlation. Finally determining the equation of a best fit relationship that best describes the cloud of data points in the data space. The defined equation can be used as a mean value for predicting the outcomes of the measurements performed (Larose, 2006; McDonald, 2015; Schoonjans, 2016). Therefore, the aims of the current research are:

1. The application and development of CITs through metabolomics multivariate analysis, for the detection and or improvement of existing biomarkers for NPC1 diseases diagnosis.

2. The application of CITs such as tri-modelling techniques in the detection of new biomarkers, and understand related pathways in NPC1 diseases diagnosis, through the metabolite profiling of biofluids datasets.

These research aims are presented as large sets of tasks; these will be delineated in the following section into more manageable detailed sets of achievable tasks that constitute the research objectives.

1.6. Research Objectives

Given the above research aims, the main research objectives may be further sub-divided into more simple and clearer measurable outcomes that are provided below:

1. To employ CITs such as the novel intelligent tri-modelling techniques (ITMTs) for the detection of new biomarkers, and the improvement of existing ones through the metabolomics multivariate analysis of NPC1 diseases diagnosis.

2. To use the scalar visualisation algorithm (SVA) for data visualisation and representation purposes.

3. To use support vector machine (SVM) transformed to the optimal support vector machine (OSVM) for data classification purposes, in the view of improving the identification of existing or new biomarkers in NPC 1 disease diagnosis.

4. To use principal component regression (PCR) models for regression analysis, establishing correlations between disease features and NPC1 disease diagnosis and progression.

5. To use the tri-ranking techniques (TRTs) as effective features ranking for biomarker detection.

6. To finally use the intelligent technology task fit model (ITTFM) to improve metabolomics multivariate analysis processes.

The aforementioned objectives highlighted allow researchers to present the main contributions made to the present research field in the next session.

1.7. Contribution to Knowledge

This research overall aimed at employing Computational Intelligence Techniques (CITs), in particular the intelligent tri-modelling techniques (ITMTs), in NMR-linked metabolomics analysis to seek and discover new biomarkers, or improve existing ones in diseases diagnosis.

In this respect, the findings of this study will become part of the literature in this research field by demonstrating that;

1. Application of CITs, especially the ITMTs has improved biomarker discovery in NPC1 diseases and the NPC1 disease-assoCITaed liver damage.

2. Applying ITMTs has led to potential new biomarker discovery via metabolomics multivariate analysis for the benefit of NPC1 diseases diagnosis.

3. Using ITMTs has allowed and improved the prediction and the correlation between NPC1 diseases features and the NPC1 disease severity and progression.

4. Applying tri-ranking techniques (TRTs) has allowed and improved biomarker detection.

5. Applying the intelligent technology task fit model (ITTFM) has improved the application of CITs in metabolomics multivariate analysis processes generally.

The above highlighted contributions were achieved based on the structure of the different chapters presented below.

1.8. Research Outline

The current thesis encompasses seven chapters establishing a logical structure for it, and the clarity in the arguments used throughout the whole research process. These chapters are outlined in detail below:

Chapter one is the introductory chapter that outlines the study background in relation to previous and recent research in the field of metabolomics multivariate analysis research, especially with respect to biomarker discovery. Research motivations and research questions were then addressed, by attempting to understand the reasons for them, whilst making connections to the different questions arising within this study. Responding to these questions guides the researcher to the successful completion of this project.

This research programme's aims and objectives were deducted from the research questions, with the second term being a more detailed and achievable outcome of the first one. Furthermore, research contributions including achievements made have been presented in terms of how they will be integrated to the current literature in this field of research. Finally, the different chapters, alongside their specific contributions, have been presented in this research outline.

Chapter two encompasses a literature review, and provides a comprehensive overview of research and literature available in the area of metabolomics multivariate analysis, especially those related to diseases diagnosis studies. It commences with some clarifications concerning the main terms related to metabolomics research. This includes metabolism, metabolites, metabolomes, etc., and the manner in which they are understood and employed in this research. Means of avoiding confusing terminologies, leading to common misunderstandings and mis-interpretations, has also been clarified. This chapter then introduces some recent research work involving metabolomics multivariate analysis, metabolite profiling, and the main CITs used in this research. Finally, research validation techniques were then reviewed, ensuring that the proposed research project follows standard validation techniques.

Chapter three describes the methodology adopted in this research, and begins with an introductory part, and then the research paradigms and research approaches involved are reviewed, with more emphasis placed on quantitative methods, given that the datasets employed and the methods utilised are exclusively quantitative ones. Moreover, the intelligent technology task fit model (ITTFM) was developed. Therefore, differing algorithms and functions were written to ease and support the current research project.

Chapter four, however, highlights the research design adopted to plan the present research as a whole, so that all the steps are coordinated. In addition, the tools involved in the data analysis stages were scrutinised and properly chosen, ensuring that they are in line with the research questions and the expected outcomes. Furthermore, the research validation models applied were visited to ascertain that they facilitate meeting the assigned objectives.

Chapter five focuses on data analysis and the results of the process stages. This is related to the different sets of results obtained from every single study conducted, and includes the results

based on existing biomarkers. New biomarkers discovered were biochemically identified, and the overall classifier performances and ranking results are provided. Data were analysed by applying various techniques, the results obtained related to the disease diagnosis, and the biomarkers detected by the employment of differing models and ranking techniques are displayed. A comparison was made in order to validate these techniques, models and findings.

Chapter six presents the discussion part of the study. The critical discussion conducted focuses on the results generated and the major findings. The suggested pathways related to the disease diagnosis were established. The pathway analysis firstly focuses on the NPC1 disease diagnosis in terms of statistical results generated, and finally the pathways detected in relation to the major biomarkers present in that pathway. Secondly, the same analysis is conducted in connection to the more severe form of the NPC disease that gives rise to liver dysfunction and damage in the case of NPC mouse model. Finally, more emphasis is placed on the biomarkers discovered.

Chapter seven is a closing chapter outlining the contribution of the present research in terms of classifier performances, and metrics improvements when compared to precedent research conducted. In addition, the new biomarkers discovered and their wider implication in disease diagnosis are highlighted. In addition, the areas where further research could be performed are outlined, ensuring that research on metabolomics multivariate analysis for biomarker discovery in NPC1 disease diagnosis is moved forward.

"*You cannot open a book without learning something"*

Confucius

2. LITERATURE REVIEW

2.1. Introduction

This Chapter reviews the most recent and relevant literature in the field of metabolomics multivariate analysis of biofluids, using Computational Intelligence Techniques (CITs). The use of computer technologies for the purpose of disease diagnosis and especially in biomarker discovery is a thrilling and very exciting research area, in which much has been conducted to improve decision-making processes for clinicians and supporting research tasks (Beger, 2013).

Thus, research on disease diagnosis process is a critical research area, mostly because it involves and directly impact on human life. Typically, when diagnosing a patient, clinicians making the wrong diagnosis may become a life-threatening issue. This is one of the main reasons why it was particularly important to use animals to measure drugs' efficacies, toxiCITies, etc., in laboratory evaluations prior to administration to humans. In this respect, the research methods and findings involving new biomarkers discoveries have to be approved through the cumbersome but necessary route of validation processes (Beger, 2013).

However, the complexity of human diseases which involve critical interactions between molecular and cellular systems, and which impact on disease development processes, require the use of different approaches. Therefore, animal and non-animal models are necessary for a wider understanding of the mode of operation of the disease, and most importantly the disease aetiology, physiopathology and pathogenesis. In addition, using experimental animals in research investigations is believed to be one of the most efficient ways of understanding the complex relationship between molecules, cells and organs with respect to disease development (Casals et al., 2011; King, 2012; Neuhaus and Halver, 2013).

Nevertheless, the detection of the main causes of the disease can be achieved through the discovery of major biomarkers. Indeed, in the research areas related to disease diagnosis, the use of biomarkers has been welcomed by researchers for a range of reasons. This includes the fact that it may assist the development of more effective and efficient treatments for the diseases under investigation, while easing the developmental steps towards and effective drug design and evaluations (Beger, 2013; Chau et al., 2008).

For the reasons noted above, reviewing other researchers' work could be equally important for this researcher, since it has enabled him to have a wider overview of what has been performed to date in this field, together with the techniques utilised in the data collection and analysis steps. It also enabled the author to learn what was achieved in terms of findings and discoveries, the differing limitations of other researchers' works, and also gaps in knowledge that would help to plan and design the present research project, via the solution of existing issues (Beger, 2013). Based on these precedents, the current literature review encompasses the following sections:

Section 1 - Chapter introduction;
Section 2 - Metabolomics and the definition of important terms such as metabolites, metabolism, etc.;
Section 3 - Computational Intelligence Techniques (CITs) and their use in the empirical work;
Section 4 - An introduction to the notion of nuclear magnetic resonance (NMR) spectroscopy and the related spectra acquisition;
Section 5 - Metabolomics multivariate analysis of biofluids, with details also provided on the Niemann-Pick Type C1 disease and its associated liver dysfunction;
Finally, Section 6 - An overall summary of the Chapter.

2.2. Metabolomics

2.2.1. Introduction

In this section the term metabolomics is defined, starting with all the related notions, including metabolites, metabolisms, metabolomes, etc. In this manner, the author has clarified how the terms were understood and used by other researchers. Below are given these series of definitions.

2.2.2. Definitions: Metabolites, Metabolism, and Metabolomes

Metabolites are considered to be the end-products of cellular regulatory processes, and as such their levels are regarded as the ultimate response of biological systems to generic environmental changes (Fiehn, 2002). They also serve as chemical entities that can be analysed using standard chemical analysis techniques such as molecular spectroscopy and mass spectroscopy (MS) (Goodacre et al., 2004). However, the use of further analysis techniques could prove to be means of improving the analysis quality; this includes coupling MS to gas chromatographic (GC) or liquid chromatographic (LC) techniques. Moreover, new techniques such as NMR spectroscopy with a very high-throughput capability could help in handling a large quantity of data, whilst allowing the extraction of a large amount of biomolecular information, especially when combined with tools such as SVM (Salazar et al., 2012).

The term ***metabolome*** was introduced several decades ago and it is not a neologism. In fact, it was used for the first time in 1998 in a paper dealing with yeast genome; all the other "omics" terms such as proteome, and transcriptome were already in use (Fiehn et al., 2002; Schripsema, 2010). In this respect, metabolomes are said to describe the complete set of all the native small molecules which are non-polymeric compounds, participating in the general metabolic reaction, with the specificity that they are required mainly for maintenance, growth and normal function of all cells. This view is compatible with these other definitions, which affirmed that metabolomes represented the complete set of all small molecules, or low-molecular-mass (< 1 kDa) components in a biological sample participating in general metabolic reactions, or the set of metabolites synthesised by an organism (Beecher, 2003; Fiehn, 2002; Kosmides et al., 2013). However, some researchers have used metabolomes to express different reality/phenomenon. For instance, Goodacre et al., (2004) affirmed that metabolomes corresponded to metabolite profiling.

As far as this research work is concerned, the term metabolomes representing a set of metabolites and is not synonymous with metabolite profiling (defined in 2.2.3.), i.e. associated with data analysis techniques. After these first series of definitions, the other terms involved in the analysis process that is described as metabolomics and metabonomics are presented below.

2.2.3. Metabolomics

- **Definition of Metabolomics**

The term metabolomics appeared in 1990 and 2000 respectively, broadly defining the study of metabolomes. The main and more comprehensive definitions of this term so far given are the ones mentioned on Xenobiotica 1999; where the term metabolomics is related to the study of the quantitative complement of metabolites in a biological system, the changes in metabolite concentrations or fluxes related to genetic or environmental perturbation. However, since then several complementary definitions have been provided in relations to this term, but the focus in this research will be on three main points that are keys to it.

1) Firstly, Nicholson et al. (2003) affirmed that metabolomics is the systematic investigation of the metabolic response of living systems to any environmental stimuli. Going deeper into this definition, researchers emphasised that metabolomics was a new research technology concerned with the systematic study of small molecules found in organisms, cells and tissues, in addition to biofluids such as saliva, urine, blood plasma, etc., (Lau et al., 2016; Shiryaeva, 2006; Sun et al., 2013). It was also suggested that large amounts of data can be generated, gathered and then analysed using various methods including mass spectroscopy (Harrigan and Goodacre, 2003; Nicholson et al., 1999; Suberu et al., 2016; Tomita and Nishioka, 2006; Vuckovic, 2012). However, NMR spectroscopy using high-throughput capabilities that required a small level of data preparation, amongst other advantages, appears to be the first choice in metabolomics research (Goodacre et al., 2004). Furthermore, Goodacre et al. (2004) argued that amongst these larger amounts of data available, only a handful will be necessary to fully describe, analyse, and understand the issue being investigated.

2) Indeed, by applying data visualisation, dimensionality reduction techniques such as feature selections, by the means of CITs, e.g. SVA and PCA in the PCR model, high-dimensional datasets are transformed to low dimensional ones to develop these models. The models (SVA, SVM, and PCR) will be used to support the thorough body fluid analyses necessary for understanding the underlying changes in the metabolomes at the precise molecular level.

3) Moreover, Fiehn (2002, 2006) suggested that metabolomics is more related to the unbiased quantitative and qualitative analysis of samples containing biochemical intermediates. This researcher added that because there was no single technology that could carry out a thorough comprehensive analysis, metabolomics was dealt with by multiple analysis techniques supported by unbiased software. Therefore, metabolomics should not be restricted to the metabolites' physiochemical properties that include the molecular mass, chemical structure, etc. Hence, tools such as scalar visualisation algorithm (SVA), support vector machines (SVM), principal components regression (PCR), etc., have been used in metabolomics research, and could even be combined in a single study for a multi-dimensional analysis of a given dataset (Zhang et al., 2014). Indeed, the use of different analytical techniques provides the opportunity to develop robust models leading to the detection of more relevant biomarkers for the research analysis ahead.

4) Finally, presented as a new powerful technology, metabolomics can be used to successfully perform biofluid-based metabolite profiling, in order to segregate between diseased and healthy control individuals (He et al., 2009; Wang et al., 2007; Zhang et al., 2014). Consequently, metabolomics has been identified as the comprehensive analytical approach that would provide complementary information compared to other omics research, such as genomics, transcriptomics, and proteomics, throughout low-molecular-mass biomolecule analysis and which is applicable to biofluids and their metabolites therein (S. Qi et al., 2012).

These major points and definitions are regarded as core to this project, and should be used for monitoring the overall research work. However, differential analysis techniques have been employed and have also been defined in the present study.

- **Metabolite Profiling**

Shulaev (2006) reported that metabolite profiling is related to the measurement of metabolite levels in a sample, in addition to a biomedical activity enabling clinicians to employ biofluids

in the assessment of patients' health conditions. Therefore, the term metabolite profiling refers to the quantitative analysis of a set of metabolites in a set of biochemical pathways, or, alternatively, a specific class of compounds.

In this regard, Fiehn (2006) stated that metabolite profiling has confined itself only in the scrutiny of a certain range, or a pre-determined number of compound classes. Therefore, as far as data analysis is concerned, a single analytical platform should be sufficient for this analysis model. Suberu et al. (2016) added that metabolic profiling provides a microscopic view of metabolite analysis, based on a targeted analysis approach using mass spectroscopy. Consequently, Bartel et al. (2013) believed that this term should be regarded as the functional signature of physiological state, affecting both genetic regulation and environmental factors.

Hence, in the current study, the term metabolic profiling is related to the quantitative analysis of a set of biochemical metabolites. Following these definitions of metabolic profiling, there remains a requirement to define the next important terms in the following metabolic fingerprinting sub-section highlighted below.

- **Metabolic Fingerprinting**

Dettmer et al. (2007) defined the term metabolic fingerprinting as a full screening approach, allowing an unbiased classification of samples based on metabolite patterns, also known as fingerprinting. In addition, the genetic perturbations that the whole biological systems undergo as a response to the transformations in a disease environment, and the possible identification of discriminating metabolites which can follow, is referred to as metabolic fingerprinting.

Based on the precedent definition, some researchers termed metabolic fingerprinting as a method allowing the discrimination between samples based on the growth condition, developmental stages, their origin, or on their biological relevance. Further, the same researchers related the term metabolomics to metabolite fingerprinting, with the former being the measurement of the latter (Goodacre et al., 2004; Kruger et al., 2008). However, some researchers questioned two main points, which were the relationship between metabolic fingerprinting and the sample origin (Mussap et al., 2013); and whether metabolite fingerprinting is related to the whole metabolomes or part of it (Dettmer et al., 2007; Scholz et al., 2004; Shulaev, 2006). As far as this research is concerned, the term metabolic fingerprinting

refers to the analysis of biological samples based on their origin; however, it does not concern the whole system metabolomes involved.

After this series of definitions, the next focus is the review of the CITs involved in the metabolomics research.

2.3. Computational Intelligence Techniques (CITs)

Different computational intelligence techniques (CITs) are available, and based on the objectives defined, different combinations of these techniques could help in answering many research questions. Below are presented the three main CITs that will be supporting this research. These include the scalar visualisation algorithm (SVA) using the technique of the 'look-up' table to convert indexes into 'colours' stored in the look-up table. On the other hand, the support vector machine (SVM) is an algorithm based on supervised learning techniques to produce a deterministic classification of the input variables. Finally, the principal components regression (PCR) strategy uses algorithms to correlate one nominal (dependent) variable which is a dichotomous or binary outcome variable in the form of 1 and 0 or diseased and healthy control, etc., and several independent or non-independent predictive variables (Schoonjans, 2016).

2.3.1. Scalar Visualisation Algorithm (SVA)

- **Principle**

The word 'visualisation' in the field of computer science, engineering, etc., relates to an area in which researchers can explore, measure and stimulate, while mining into data in order to get insight into the underlying relationship within the data structures (Johnsona, 2012). Three main visualisation algorithms exist, including the scalar algorithm, the vector algorithm and the tensor algorithm. The main focus in this study is on the first algorithm mentioned. The scalar visualisation algorithm (SVA) is used in producing images of the different diseases datasets showing the importance and interconnection between features in disease diagnosis. Using the colour mapping technique to display features in such a manner that changes in the features' structures can be displayed and viewed. The same researcher supports that the visualisation

deals with generating images that convey salient information, while improving our understanding of underlying processes. Therefore, Johnsona (2012) stated that data visualisation carries such important value in terms of the understanding of the large amount of data and information flowing in various fields of research, including biomedical and medical science, bio-informatics, pharmacy, particle physics, engineering, etc.

The scalar visualisation algorithm is based on the principle of creating a relationship between the spectrum peak value considered as index and the colour look-up table. In this manner, a spectrum peak will correspond to a colour and the underlying changes at molecular level can be visualised by changes in the colour in the map hence generated. SVA changes the data structure, like changing or mapping the scalar data into colour. Subsequently, some applications of the scalar visualisation algorithm are presented.

- **Model Applications**

The scalar visualisation algorithm has been applied in different field of research. In the following research, the model has been applied to map the additional scalar values to the output in order to make the contrast more apparent. Indeed, some data type carries too many different kind of information in such a way that even colour is not enough to represent them all in the output. The image process and the texture mapping are combined to map scalar values to the local image contrast (Pan et al., 2004). In this other study cerebral spatial structure of the cerebral vessels were visualised in three dimensions allowing an effective disease diagnosis. The combination of distance colour blending, a spectroscopic technique and the CUDA-based volume representation and the related transfer function have helped the team of researchers to make it possible for surgical team to observe from different angles the vascular areas. This technique has enabled to improve the cerebrovascular structure, while revealing adjacent areas that can make a massive difference during surgical intervention (Luo, 2013).

In addition to visualising the disease features relationships, segregating diseased group from healthy control individuals is another way of having more insight into the disease underlying structure. The support vector machine is a classification algorithm that can help in this respect and is next studied in this research.

2.3.2. Support Vector Machine (SVM)

- **Principle**

Support vector machine (SVM) is a classification algorithm based on a heuristic process, allowing the development of a prediction method. As a classifier, SVM was developed in the 1990's by Vapnik and his research team. The rationale behind this method is to find the best fit hyperplane allowing through input variables a separation between two different categories. This could be used for example on a binary classification that requires the segregation between diseased and healthy control individuals in a disease dataset. However, when this is not possible, a kernel function φ can be applied that transforms the input variables, such as in increasing the dimensionality of the input space, to find a separation boundary between the two classes (Salazar et al., 2012). Thus, by applying φ to the original dataset D_0, a new dataset is generated D_i, such that:

$$\forall\, x_i \in D_0, \exists \phi / \phi : x_i \;\longmapsto\; y_i = \phi\,(x_i) : (\mathbf{1})$$

$$Hence, any\ x_i\ belonging\ to\ D_0\ there\ is\ \phi\ such\ that\ y_i = \phi\,(x_i).\ Further,$$

$$\{(\Phi(x_i), y_i)\}_{i=1}^{n}\ with\ y_i = \{1,0\}, indicating\ binary\ classification, therefore$$

$$the\ closest\ hyperplane\ to\ the\ data\ points\ from\ each\ of\ the\ two\ classes, has\ the$$

$$equation: W^T \phi\,(x) + b = 0 : (2')\ that\ is\ linear, and\ linear\ separability\ involves$$

$$\exists\, W, b\ /\ W^T \phi(x) + b = 1 : (\mathbf{2})$$ (Salazar et al., 2012; Verplancke et al., 2008).

$$By\ applying\ this\ assumption, it\ is\ derived\ that:$$

$$W^T \phi\,(x) + b \begin{cases} \geq 1, & if\ y_i = 1 \\ \leq -1, & if\ y_i = -1 \end{cases},\ i = 1, 2, \ldots., n : (\mathbf{3})$$

$$The\ distance\ separating\ the\ two\ classes, is\ the\ double\ of\ the\ one\ between\ a\ class$$

$and\ the\ hyperplane, and\ is\colon\ \frac{2}{\|w\|}. Therefore\ maximising\ the\ margin\ is\ equivalent\colon$

$to\ solving\colon\ \min_{w,b}\|w\|^2\ , with\ y_i\ (W^T\phi\ (x)+b)\geq 1\ ,\qquad for\ \ i=1,2,\ldots\ldots,n:(\mathbf{4})$

$If\ \ w^*\ and\ \ b^*\ represent\ a\ solution\ to\ \ equation\ \ (4), the\ hyperplane\ equation\colon$

$becomes\colon D^*(x)=(W^*)^T\ \phi\ (x\ )+b^*)=0.\ \ All\ the\ values\ \phi\ (x_i)\ satisfying$

$equation\ (4)\ are\ known\ as\ a\ support\ vectors.\ \ Hence, SVM\ as\ deterministic$

$classifier\ can\ be\ a\ powerful\ tool\ for\ extracting\ useful\ information\ from$

$large\ dataset, especially\ when\ there\ are\ more\ variables\ than\ observations$

(Salazar et al., 2012; Verplancke et al., 2008).

- **Model Applications**

The support vector machine (SVM) serves as a deterministic classification algorithm using a heuristic method for prediction purposes. SVM is based on the principle of detecting a hyperplane that is used as a separation n-dimensional plane between two different classes (Verplancke et al., 2008). The process of defining a hyperplane can be achieved through a straightforward process or through a transformation using a linear kernel function, especially when it is not possible to find such hyperplane in a lower dimensional input space, by mapping the dataset into a higher dimensional features space. In this manner, such a transformation can allow the detection of the separation hyperplane (Grootveld, 2014; Salazar et al., 2012; Zhang et al., 2006).

In the specific case of binary classifications, the SVM has been employed in order to discriminate between diseased individuals and healthy control ones in relation to two biomedical datasets, i.e. the NPC1 blood plasma and liver dysfunction disease datasets.

However, in this research project, the SVM approach applied employed the recursive features elimination technique to remove the less significant features in this binary classification process (Guyon et al., 2002), combined with the k folds stratified cross-validation technique, in which the dataset is divided into a total of 5, 10 or 15 groups or folds. During the modelling process, one fold is purposely omitted, also known as leave-one-out, and this process is repeated n times. The technique is used to validate the whole modelling process (Grootveld, 2014). In this research, the classification of the patients into two groups is insufficient, since one of the main aims is to connect the diseases features to the disease progression. A regression analysis would help in this regard, and therefore the principal components regression strategy described below was applied.

2.3.3. Principal Component Regression (PCR)

- **The Principle**

The technique of principal components regression (PCR) is mainly the regression analysis of Y upon a series of principal components, and has been mainly used to address collinearity issues in the dataset, especially amongst variables or predictors (Cook, 2007; Massy, 1965). Researchers support that when many variables are potential explanatory ones, it becomes more complicated to select the variable(s) responsible for the variability observed in the dataset examined (Massy, 1965; Puelz et al., 2017). However, the conversion of these potential explanatory variables to one or several main explanatory variable(s) reduces the variable space, a process easing the explanation and the overall understanding of the modifications sought. Finally, the nominal dependent variable can be regressed upon the principal components. Hence, when the inter-relations between variables are not clearly understood, and when too many variables are candidates for the variability explanation, then it becomes necessary to use the principal components for more visibility in the explanatory variability of the disease diagnosis and progression. PCR in this research has been applied as a combination of the principal components analysis (PCA) and the multiple logistic regression (MLR) approach.

- **Principal Component Analysis (PCA)**

- **Principles**

PCA is a statistical procedure developed in 1901 by Karl Pearson, who based his technique on the famous theorem of the principal axis in geometry, related to the major and minor axes of a hyperbola (for generality). It is considered as one of the oldest technique for reducing the data dimensionality in the area of multivariate data analysis (Cook, 2007).

As an exploratory method, principal components analysis involves orthogonal transformation of the original features into a new reference of features, also known as factors or principal components (PCs) that are a linear combination of the original features (Anton, 1987; Manfredi, 2013). In view of the fact that the PCs are orthogonal to each other, they therefore carry independent information. The first principal component accounts for the main variance in the original dataset, which corresponds to the main independent information; however, the other principal components carry the residual information (Manfredi, 2013).

PCA provides important information on the principal axes for data analysis, including the factor scores and loadings. The first corresponding to the sample coordinates in the principal component reference, whilst the second gives the coefficients of the linear combination of the original features on a given PC. The gathered information can be interpreted with regards to the potential and most relevant biomarkers present in such datasets (Manfredi, 2013). Some applications of principal component analysis are examined below.

- **Applications of PCA**

Different cases have been reported in the literature in relation to the application of PCA as a visualisation tool to determine whether there is total separation between classes in the new features that is the PCs' space. In respect to the understanding of the likelihood of a patient developing diseases such as NPC1 disease, PCA can provide very useful information for the detection of candidate biomarkers (Manfredi, 2013).

PCA was employed for the identification of blood plasma biomarkers for NPC1 disease. Indeed, this study aimed to identify the pathological mechanisms that underlines NPC1 disease, in order to detect biomarkers. In this regard, the change in liver expression in the NPC1 mouse model at different steps of disease progression was also studied. PCA was able to identify three different expression groups, including 1-week-old control and mutant mice, 3- to 11-week-old *Npc1*$^{-/-}$ mice, and the 3- to 11-week-old control mice (Cluzeau et al., 2012). Therefore, the use

of statistical techniques such as PCA in the interpretation of metabolomic, genomic, etc., datasets allows the extraction of biomarker(s) that can inform us in disease diagnosis and prognosis (Wang, 2008).

One example of the applications of PCA is another "omics" strategy focused on the potential effects of pollutants on fish. Fish plasma samples were collected and analysed using a surface enhanced laser desorption ionization – time of flight mass spectroscopy (SELDI-TOF MS) that is a gel-free proteomic technique. Hence, the protein expression in fish and the transformations at organism level were investigated (Nilsen et al., 2011). The results showed a strong correlation between the first principal component PC1 (explaining 23.6% variability) and the gonadosomatic index (GSI) value. Further analysis of PC2 and PC3 explained respectively 18.6% and 9.1% variability respectively, and highlighted an additional relationship related to alkylphenol exposure. This, in turn, confirmed the ability of the PCA to distinguish between females exposed to alkylphenols from those unexposed, which could disclose potential biomarker candidates specific to such exposure (Nilsen et al., 2011).

Principal component analysis can be used to separate different groups of individuals or patients in relation to disease studies. Additionally, it can be combined with other algorithms while trying to understand features involved in disease progression. This includes the multiple logistic regression (MLR) approach, which is described in the sub-section below.

Multiple Logistic Regression (MLR)

- **The Principle**

Schoonjans, (2016) defines multiple logistic regression (MLR) as a statistic model used in data analysis, which features one dependent nominal variable that has two values. In such cases, the dependent variable is a dichotomous or binary coded value of 1 for (diseased, male, yes, etc.), and 0 for (healthy, female, no, etc.), and two or more measurements which could be independent, or non-independent and correlated. He added that MLR allows use of variation in the measurements to predict the value of the dependent variable.

Indeed, he also noted that applying MLR aimed at finding the best fitting model describing the relationship between the dichotomous characteristic of interest (dependent nominal variable,

i.e. an outcome variable), and a set of independent predictors or explanatory variables. The probability of predicting an outcome is given by the formula:

$$Y' = Y_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 \ldots + \alpha_n X_n : (\mathbf{5})$$

$Higgins, (2005) showed\ that\ Y_0\ intercept\ of\ Y', and\ the\ values\ \alpha_i\ the\ change\ in$

$Y'\ for\ each\ 1\ increment\ change\ in\ X_i\ can\ be\ determined.\ For\ the\ calculation,$

$refer\ to\ (Higgins, 2005).$

Moreover, the logistic regression generates the regression coefficients and their associated standard errors, with a formula to predict the logit transformation of the probability of the presence of the response or the outcome as mentioned below:

$$logit\ (p) = Y' = Y_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 \ldots + \alpha_n X_n : (\mathbf{6}),$$

$where\ p\ is\ the\ probability\ of\ presence\ of\ a\ given\ outcome.$

$The\ logit\ transformation\ is\ defined\ as\ the\ logged\ of\ the\ odds\ defined\ as\ follow:$

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ characteristic}{probability\ of\ absence\ of\ characteristic}$$

$$logit\ (p) = ln\ \frac{p}{1-p}$$ (Schoonjans, 2016).

The estimation in logistic regression chooses parameters that maximise the likelihood of observing the sample values, rather than choosing the parameters that minimise the sum of square errors (Higgins, 2005; Schoonjans, 2016). Managing high-dimensional variables with only a few samples is highly complex in regression analysis; however, regressing Y on the principal components seems to be effective, providing acceptable results which are explored further here through principal components regression (PCR) (Artemiou and Li, 2009; Li, and Chiaromonte, 2007).

- **Applications of the MLR Model**

Different applications of MLR have been reported to predict the likelihood of a patient developing a given disease (Abbott and Carroll, 1984; Schoonjans, 2016; Taooka et al., 2014). In a biomedical research study, the binary nominal dependent variable Y will take the binary values corresponding to either diseased or healthy individuals. Similarly, MLR is used in the current thesis to predict the likelihood of an individual being correctly identified as an NPC1 disease patient, which corresponds to the individual being likely to develop the disease at an early stage or being at risk of sudden death (Taooka et al., 2014).

Equally, MLR has been used in the analysis of prolonged QTc interval in the treatment of pneumonia patients. In the heart electrical cycle related to the ventricle depolarisation and repolarisation, the QTc corresponds to the time elapsed between the start of the Q wave and the end of the T wave. The datasets analysed in the record of clinical and laboratory data, and contained information such as ECG findings, blood chemistry data, etc. The multiple logistic regression (MLR) model was used to compare the datasets (Taooka et al., 2014).

The investigation of QTc as one of the risk factors in elderly pneumonia patients was performed by a team of researchers. The results acquired showed that the prolonged interval of the heart depolarisation and repolarisation, that represents a risk factor for ventricular dysrhythmia (causing sudden death) of QTc > 0.44 seconds is a potential prognosis factor for pneumonia in elderly patients. In fact, MLR analysis showed that the QTc prolonged interval also featured as a predictive factor of increased mortality in pneumonia patients (Taooka et al., 2014; World Wide Antimalarial Resistance Network, 2012).

Following the presentation of certain applications of MLR, and with the two steps (i.e. PCA and MLR) combined together, the principal component regression model has hence been further developed. Some applications of the PCR model are provided below.

- **PCR Model Applications**

Several applications of PCR model have been mentioned in different literature. For example, PCR was utilised to understand the traits variance related to underlying genetic transformations. The objective of the study was to assess whether or not the multiple correlation

on single nucleotide polymorphisms (SNPs) can be detrimental to the selection of a given gene as main biomarkers for trait variance (Shen and Zhu, 2009). Similar studies have shown the importance of the interaction between different loci within a gene, which can generate more information and finally improve the detection power (Schaid et al, 2002).

However, determination of the number of PCs upon which to regress the dependent variable Y' is not rationally established, but the common suggestion and applied technique is to use several PCs among the first that account for 80 to 90 % of the variability, also considering that these may just not be strongly related or even unrelated to the outcome (Gauderman et al., 2007; Wang and Abbott, 2008). Nevertheless, the author of this thesis believes that it is worth expanding the variability space in an attempt to involve more PCs in the understanding of the phenomenon under investigation. Indeed, involving more principal components in the explanation of the maximum variability within the dataset (80% and 90% variability), is more likely to traduce the whole reality of the changes happening at genes level.

The combination PCA and logistic regression was applied to develop a model that will facilitate the construction of a valid coronary heart disease risk score in comparison to the traditional clinical risk one. For this purpose, the researchers generated the genetics risk scores (GRS) of a cohort of 49,310 SNPs based on the cardiogram meta-analysis of coronary heart disease (CHD) of the cohort of individuals recruited to the study. Different tests were performed on 5 different types of population at risk. Clinical scores such as the Framingham risk score (FRS) was also considered in this research. The results showed that integrating the GSR and FRS values improved the risk prediction for those aged 60 years, and plus for more than 10 years. Notwithstanding, the GRS detected different risk patterns with men being more at risk, which was between 12 to 18 years earlier than those less at risk (Abraham et al., 2016).

These studies highlight the importance of the application of PCR to correlate disease features and the diagnostic. Nevertheless, metabolomics multivariate analysis implies also the use of ^{1}H NMR spectroscopic technique that has been used to convert the blood and liver extract samples into dataset, and the technique is next highlighted.

2.4. Nuclear Magnetic Resonance (NMR) and Data Acquisition

2.4.1. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is now a widely used analytical technique with a wide variety of research applications, including pharmaceutical and biotechnology analysis, clinical research, etc. The technique is also applicable to physicians' use as magnetic resonance imaging (MRI), since the principle is the same (Amathieu, 2016; Bruker Corporation, 2013). NMR spectroscopic analysis of biofluids, however, it is a more reproducible technique that requires little sample preparation, is able to provide a fast, comprehensive information regarding the biomolecular composition of metabolites therein, together with a thorough and full structural elucidation (Joshi, 2012).

2.4.2. Theory and Principles

The main concept of nuclear magnetic resonance is based on the phenomenon of nuclear spin. Indeed, atoms in molecules have discrete energy levels characterised by electronic, vibrational, and rotational or spin states. When placed in an electromagnetic field (B_0) generating electromagnetic radiation, atoms present in chemical compounds will absorb and emit photons, in such a manner that the energy of the photon will correspond to one of the energy levels of the atoms. The different types of spectroscopy are linked to the frequency involved since the energy level is proportional to the frequency according to the equation:

$$E = h\mu : (\mathbf{7});\ h = 6.62\ .10^{-34}\ m^2\ kg\ s^{-1}\ \ is\ the\ planck\ constant;\ \mu\ the\ frequency$$

NMR spectra are obtained throughout a range of frequencies (10-800 MHz), which correspond to the resonance frequencies of certain spectrometers. Therefore, NMR is viewed as the study of the magnetic properties and the energy of atomic nuclei when placed in a powerful external magnetic field, B. Thus, when a nucleus of mass m and charge q = e is rotating with a constant veloCITy v at a constant distance r from the rotational axis, the momentum μ can be expressed by:

$$\mu = \left[\frac{ev}{2\pi r}\right]\pi r^2 : (10).\ \ Additionally\ It\ can\ be\ established\ that:$$

$$\mu = mvr = \left(\frac{e}{2m}\right) L = \gamma . L : (\mathbf{8})$$

A rotating nucleus or object placed in a magnetic field B_0, *will be exposed to a torque*: $T = \mu B_0$, *with* $\left(\frac{d\mu}{dt}\right) = \gamma\mu . B_0 = \mu . \gamma B_0 : (\mathbf{9})$

A precession of μ *about* B_0 *can be described by* $\left(\frac{d\mu}{dt}\right) = \mu . \omega_0 : (\mathbf{10})$

The Larmor equation establishes that: $\omega_0 = \gamma B_0 : (\mathbf{11})$ (De Graaf, 2007).

Further information regarding the theory and principles of the ^{1}H NMR spectroscopic technique can be found in (De Graaf, 2007).

2.4.3. Generation on NMR Free Induction Decay (FID) File

When biofluids or tissue samples extracts are subjected to high-resolution NMR analysis, an FID file containing all the information in the form of excitation frequency or amplitude of signal emitted is provided. Thus, when the RF emitted by the magnetic field is equal to the frequency of the nucleus, there is absorption followed by an emission of a signal. The frequency (ppm, known as chemical shift) values for each nucleus studied are critically dependent on its magnetic and therefore chemical environment within a particular molecule. Therefore, the FID file contains information related to the excitation amplitude given the ppm value of the chemical shift. Finally, the raw spectra information contained in the FID files are converted into Excel files, before they undergo further processing, including baseline correction, zero-filling, Fourier-Transform, etc.

2.4.4. Zero-Filling

The technique of zero-filling has the objective of improving the spectral resolution of the NMR spectra generated. However, spectral resolution is usually too low; hence, resolving the problem is more related to knowledge of the spectral amplitude and the intermediate frequencies. De Graaf (2007) argued that decreasing spectral width or increasing the acquisition time appears to serve as the best options for solving this issue. However, the latter solution increases the data storage, resulting in an increase of the noise contribution. Furthermore, extending the acquired FID automatically increases the acquisition time, a process solving the spectral resolution problem. Therefore, prior to performing a Fourier

Transformation, a string of points of zero amplitude are added to the FID files, a technique known as zero-filling.

2.4.5. Fourier-Transformation Process

During Fourier Transformation, the RF excitation creates a complete excitation state of the nucleus, while the magnetisation is precessing about B_0 and inducing an electromagnetic field (emf) within the receiving coil. Additionally, spins can exchange energy between themselves, hence the spin-spin relaxation time T_2 causes the emf to behave like a decreasing function of time. However, a local distribution of B_0 is created across samples, followed by a subsequent distribution of the Larmor frequencies. The frequency distribution causes a rapid loss of magnetisation, which is more profound than the loss of magnetisation experienced/sustained during the T_2 relaxation time (De Graaf, 2007). The Larmor Frequency for proton spin is given by the formula:

$$\omega_{proton\ spin} = \frac{2\mu_p B}{\hbar} \quad (\mathbf{12})$$

Therefore, the signal generated in an inhomogeneous magnetic field for a sample, with uniform proton density and a relaxation time T_2 given by:

$$M_{xy}(t) = M_{xy}(0)e^{-\frac{t}{T_2}} \int_r^{\infty} e^{+i\gamma \Delta B_0(r)t}\, dr = M_{xy}(0)e^{-\frac{t}{T^*_2}},$$ *with* ΔB_0 *being indicative of* B_0 *inhomogeneity and* $\Delta B_0 = B_0(r) - \Delta B_{0,nom}$ *; with* $B_0(r)$ *strength of the magnetic field at position* r*, and* $B_{0,nom}$ *the nominal magnetic field strength.*

The electromagnetic signal sent to the receiver in the transverse plane precessing at the Larmor frequency, and exponentially decaying at a characteristic time constant T^*_2, is represented as a complex function:

$M_{xy}(t)$, *expressed as*:

$M_{xy}(t) = M_x(t) + iM_y(t)$; *complex transverse magnetisation with* M_x *real part, and* M_y *imaginary part, and expressed as function of the time* t.

$$M_x\,(t) = M_0 \cos[(w_0 - w)t + \emptyset]e^{-\frac{t}{T*_2}} : (\mathbf{13})$$

$$M_y\,(t) = M_0 \sin[(w_0 - w)t + \emptyset]e^{-\frac{t}{T*_2}} : (\mathbf{14})\ with\ \emptyset\ the\ phase\ at\ time\ origin\ t = 0$$

NMR is therefore able to detect the x and y components of this complex motion. The time-dependent emf (signal amplitude) is known as free induction decay (FID). M_x (t) and M_y (t) are known as real and imaginary FIDs respectively, with sinusoidal representations. Further, M_x (t) and M_y (t) correspond to the spectra acquired during NMR spectroscopy (De Graaf, 2007). Indeed, although containing the most relevant information on the nuclear spin, the time-domain data (FID) is converted into a frequency domain (spectra) using the Fourier Transform function, which is provided below;

$$\mathrm{f(t)} \longrightarrow \mathrm{F(\omega)},\ \text{transforming FID} \longrightarrow spectra$$

$$F(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}\,dt : (\mathbf{15})$$

or

$$F(\nu) = \int_{-\infty}^{+\infty} f(t)e^{-i2\pi\nu t}\,dt : (\mathbf{16})$$

Based on the Fourier Transform, the values of the complex transverse magnetisation function can be determined (De Graaf, 2007).

2.4.6. Chemical Shift

The notion of chemical shift is applied based on the assumption that the resonance frequency ω depends on the magnetic field B_0, the gyromagnetic ratio γ, and the chemical environment of the nuclei studied. This latter effect is known as chemical shift. Indeed, nuclei are shielded by electrons surrounding them to the external magnetic field B_0. Therefore, the effect of these electrons will be to reduce the overall effect of the external magnetic field received by every single nucleus. The resulting intensity of the magnetic field B received by a given nucleus is given by the formula below, where B_0 is the external magnetic field value, and σ is the shielding constant.

$B = (1 - \sigma)B_0 : (\mathbf{17})$ (De Graaf, 2007).

Data acquired in this manner finally go through the data processing stages, in order to remove noise prior to the analysis stage. Some applications of the NMR spectroscopy to the analysis of metabolomics biofluids are highlighted below.

2.5. Metabolomics Multivariate Analysis of Biofluids – Application to the Study of Niemann-Pick Disease Type C1 (NPC1)

2.5.1. Introduction

Major research publications focused on metabolomics research related to NPC1 disease are investigated in this study. The manner in which the research was carried out, what techniques were used, what the findings were, and finally how to relate all these different steps to the current research questions. To answer each of these different questions, the following are included in this subsection: (1) an introductory component presenting the importance and the scope of this review; (2) an overview of Niemann-Pick type C1 disease; and finally (3) metabolomics multivariate analysis of the disease is discussed, with the underlying pathways related to the disease progression, together with research limitations and the gap(s) in knowledge, which are all scrutinised.

2.5.2. Niemann-Pick Diseases Types C1 and C2

Niemann-Pick disease type C is a rare neurodegenerative lysosomal storage disorder which arises from functional losses of either NPC1 or NPC2 genes. Indeed, the NPC disease is caused by the mutation of NPC1 (95%) and NPC2 genes (5%) (Vanier, 2010). This altered storage at lysosomal and endosomal levels, coupled with the cellular trafficking of cholesterol and sphingolipids to degradation, causes unesterified cholesterol accumulation in the lysosomes and late endosomes, and 80% of the patients experiencing pronounced abnormalities. In infant years and childhood, the neurological symptoms vary from delay in the sensorimotor developmental milestone, followed by falls, clumsiness, ataxia, psychiatric disturbances, etc.,

(Vanier, 2010). Although to date there is no fully effective treatment for NPC1 disease, miglustat (MGS) is the only alternative therapy approved by the Food and Drug Administration (FDA) in the USA (Fan et al., 2013). ^{1}H NMR-based metabolomics could indeed provide insights into NPC1 disease prognosis and diagnosis while presenting a clear picture of the underlying metabolic alterations in this disease's early stages, and this analysis is reviewed below.

2.5.3. Metabolomics Multivariate Analysis of NPC1 Disease

Metabolomics analysis and metabolite profiling of the NPC1 diseases, and the more severe form of the disease related to the liver dysfunction are being investigated in order to achieve deep insights at the biomolecular level. This, in turn, can help us to attain a clear understanding of the disease's aetiology. This scrutiny will provide readers with a clear picture of the changes occuring at molecular level, and the related pathways associated with disease progression.

In this regard, a team of researchers conducted a study on NPC1 disease in order to detect biomarkers that will help to clarify disease progression. Their main objective was to detect biomarkers that would provide valuable information regarding defects in metabolic pathways, information of much relevance to the development of effective treatments for NPC1 disease (Fan et al., 2013).

Their work focused on studying lipid storage at a lysosomal level in mouse tissue, and human cerebrospinal fluid, to which they applied targeted metabolomics profiling. They found that the following lipids, including sphingoid bases, monohexosylceramides (MC), and the GM2 gangliosides were highly concentrated in the liver. Additionally, high levels of GM2 and GM3 gangliosides were present in the brain (Fan et al., 2013). They decided, with the help of an animal cohort, to treat this high lysosomal concentration of sphingolipids with miglustat and 2-hydroxypropyl- β-cyclodextrin (HP- β -CD). The results of both treatments showed a decrease of the sphingolipids at lysosome level. Based on the success of this animal-based study, the same researchers extended their technique, but this time by applying similar treatments to human subjects.

They also analysed blood plasma and cerebrospinal fluid (CSF) of NPC1 patients through targeted metabolomics profiling. They discovered that biomarkers in plasma and CSF samples

responded to the therapeutic intervention. They applied MGS as drug in the treatment of the NPC1 disease and obtained a significante reduction of the level of the GM1 and GM3 gangliosides in the NPC1 subjects blood plasma. However, CSF increased MCs sphingolipids. This highlighted the fact that the sphingolipidosis during NPC1 disease was attributable to the trapping of unesterified cholesterol in the lysosome, which in turn does not permit recycling of the sphingolipids. Moreover, MGS treatment in NPC1 patients provoked profound alterations in plasma and cerebrospinal fluid sphingolipids. HP-β–CD treatment caused a similar alteration in CSF to that observed in NPC1 mouse cohorts. Furthermore, researchers noted a reduction in lipid content in liver, spleen and brain tissue after single or combined therapy of MGS with HP-β–CD. These researchers also discovered that HP-β–CD, and combined treatments were equally effective, but more effective in reducing lipid storage than MGS administered alone; this signified that HP-β–CD treatment was more effective and efficient than the MGS alone.

(Fan et al., 2013) postulated that NPC1 gene mutation was causing a severe perturbation such as intercellular cholesterol trafficking, a process resulting in a huge and uncontrolled accumulation of lipids within the lysosome. The temporary solution was the application of a therapy that provided some relief in terms of controlling the excess of lipid accumulation. Finally, the FDA in the US approved the therapies based on MGS, the glycosphingolipid inhibitor that gained approval for the treatment of NPC1 disease.

In another research investigation, Yang, (2005) showed that sterol homoeostasis dysfunction caused cholesterol accumulation, and the provocation of neuronal apoptosis was one of the main causes of NPC1 disease. Typically, blockage of cholesterol transport between the lysosome and the sterol regulatory machinery was responsible for the cholesterol homoeostasis defect.

The same researcher further investigated this disease, and followed five different Chinese patients and their family. The neurological manifestation of the disease encompassed the following signs, amongst others, specifically cerebral atrophy with psychomotor retardation, speech, and mental function deterioration, etc. Indeed, it was shown that there was a variation in the age at which the first disease symptoms appeared, and the later the appearance, the slower the disease progression. Moreover, the earlier the disease symptoms manifested, the more severe the phenotype was.

However, the validation of these findings was insufficient. Therefore, its acceptance was weakened in view of several reasons, including the limited number of NPC1 subjects, combined with the fact that the reliance on the clinical trial outcomes indicated that the long-term observation of patients, and the difficulties associated with the measurement of variance in outcomes. These reasons were sufficient to affect the therapies' reliabilities. Consequently, the FDA was sceptical about the findings and related techniques, despite being one of the rare therapies currently applied. This leaves the door open to more effective NPC1 treatment regimens to be developed.

Typically, the determination of NPC1 disease biomarkers allow a non-invasive, non-destructive, quantification of disease progression, and this development will be welcomed by all stakeholders involved in this orphan disease research. Moreover, researchers affirmed that oxysterols have been already discovered as specific biomarker for NPC1 disease; therefore, the discovery of more powerful and discriminatory biomarkers is required (Fan et al., 2013).

Following their study and discoveries, (Fan et al., 2013) suggested that it was valuable to use sphingoid base monohexosylceramides (MCs), i.e. the GM2 and GM3 gangliosides, as biomarkers for determining treatment efficacy, especially during clinical trials in monitoring NPC1 disease progression based on drug administration. They added that the sphingolipids detected fulfilled the required biomarkers discovery criteria. This included ease of detectability in CSF, and their quantification through MS/MS, etc.

However, Yang (2005) believed that in view of NPC1 disease pathogenesis complexity, including proteostasis, liquid trafficking, inflammatory, oxidative stress, and pro-apoptotic pathways, it is unlikely that a biomarker or class of biomarkers will be found that would help to define a quantifiable measure for the diagnosis, monitoring, and assessment of responses to intervention in NPC1 disease. Moreover, the latter researcher sustained that the main problem related to cholesterol homoeostasis defect triggers neuronal apoptosis. However, he added that NPC1 gene mutation and the different functional implications remained unknown (Bi and Liao, 2010, Wang, 2011, Yang, 2005).

Despite the fact that these researchers, main focus was based on the finding of new biomarkers for NPC1 disease diagnosis, they proposed further research towards the development of techniques to determine whether the altered sphingolipids in plasma are only related to NPC1

or to all other lysosomal storage diseases (Fan et al., 2013a). Thus, these researchers gave another orientation to the research, potentially ascribable to several explanations. Nevertheless, this research is keeping the current orientation with the option of finding new biomarkers in order to improve NPC1 disease and the NPC-associated liver disease treatment regimens.

After scrutinising the research on NPC1 disease, and even finding gaps in knowledge, the focus of this research area can be placed on other types of NPC1-related diseases that are investigated using metabolomics and metabolite profiling. One major aspect of this work was focused on liver dysfunction associated with this disease, as well as on the research methods applied, the data analysis techniques used and their relevance to the research questions.

In this regard, the following study concerning liver disease revealed that blood serum metabolomics could be used as a research analysis technique to differentiate between patients with liver cirrhosis, and that the application of multiple classification methods could ultimately separate patients with low, mild or severe chronic liver failure (CLF). However, in this particular case, the research was conducted by applying metabolomics using ^{1}H NMR analysis for the identification and quantification of metabolites. The detection of glutamine could help in this study for the prediction of an unfavourable outcome for patients with fulminant CLF (Amathieu, 2016).

In another study, Qi et al. (2012) focused their work on serum metabolite profiling of individuals with liver cirrhosis (LC), especially when it reaches the advanced stage of liver failure, where the liver is unable to compensate changes related to bridging fibrosis. The above researchers believed that detecting this disease in its early stages could serve to provide patients with the necessary care and also to make accurate decisions regarding the treatment to be administered, whilst ensuring the detection of efficient biomarkers related to LC.

Furthermore, the above researchers demonstrated the importance of hepatitis B virus (HBV), with a higher impact on patients developing LC. For this reason, their study was reoriented toward the analysis of the HBV-liver cirrhosis patients. They discovered that the distinction between patients in the early stage and those at the later stage of the disease could clearly be established. Furthering their study, they also found that serum metabolites could be used to achieve a features discrimination between alcoholic cirrhosis and HBV-liver cirrhosis type,

with the advantage of detecting the molecules responsible for the difference; including creatine, acetoacetate, glutamine and glutamate viewed as significant biomarkers.

This shows the enormous potential offered by metabolomics in the study of human diseases. Indeed, the use of CITs provides newly, rich and diverse research and investigation platforms in relation to enhancing our understanding of disease aetiology and diagnosis, and other opportunities to follow (Duarte et al., 2014). Nonetheless, research related to liver dysfunction disease has not yet detected relevant biomarkers that will help in the diagnosis of the disease advancement. Therefore, there is more work to be done on NPC1 disease diagnosis, in order to provide practitioners with new biomarkers that will be a breakthrough in its diagnosis and therapeutic monitoring.

2.6. Chapter Summary

This chapter provided the author with an opportunity to review the state-of-the-art of metabolomics multivariate analysis in disease diagnosis. Throughout this journey, several terms were reviewed, in such a manner that differing interpretations and understandings were unified and clarified. Moreover, different CITs were visited to give a clear picture of how they could be supporting the present research project in order to meet the research aims and objectives. Furthermore, NMR theory and principle were investigated to provide a wider understanding of the phenomenon behind it, including the way data is generated, especially NMR spectral profiles used to explain the underlying transformation at the molecular level with respect to disturbances in metabolism.

The next chapter, the research methodology section will focus on the methods employed.

"*I cannot articulate enough to express my dislike to people who think that understanding spoils your experience… How would they know*?"

Marvin Minsky

3. RESEARCH METHODOLOGY

3.1. Introduction

Methodology is related to learning and adopting several common approaches involved in researching, and it usually leads to designing the research and finishing with data gathering and data analysis. Moreover, it is a form of thinking impeding on the design and all the changes in the research organisational plan (Jonker and Pennink, 2009). Whilst other considered it as the strategy or plan of action linking the methods to the research outcomes (Creswell, 2003).

However, different combinations can be applied when it becomes a methodological choice. Depending on the type of research carried out, researchers chose the research approach that best suit their work in order to meet the research aims and objectives (Sarantakos, 2005; Wahyuni, 2012). In the present thesis, a quantitative research approach was chosen, while the intelligent technology task fit model was developed to support the whole research project.

This chapter is structured as follows. Section (1) contains the introduction with an overview of the chapter, whereas section (2) highlights the research methodology, which describes the approach, and also the theory selected. In section (3), the research method is visited, and in section (4) the samples and feature selection techniques are respectively investigated. In section (5), the research project challenges are identified. In section (6), the possible outcomes of the research work conducted are forecasted, and finally in section (7) a summary of the chapter is provided for conclusion purposes.

3.2. Research Methodology

In this research methodology section, a brief description and justification of the choice of quantitative research approach used to support this metabolomics multivariate analysis research is included. This is followed by the presentation of the intelligent technology task fit model (ITTFM) developed as an extension of the task technology fit theory (TTFT) by Goodhue and Thompson (1995), followed by a highlight of the reasons for this choice.

3.2.1. Quantitative Research Approach

- **Quantitative Research Approach Notion**

There are two main research approaches, including quantitative and qualitative research approaches. Different opinions exist regarding the use of these research approaches in science in general. Some are of the opinion that both approaches should be considered and applied together, while others take the view that they are quite different and as such should not be mixed, but rather should be considered separately. A third approach is that involving the combination of these two main approaches. This is referred to as the mixed research approach. Both approaches can utilise knowledge, techniques, etc., from each other, since the two main approaches are useful and equally valid for research purposes (Bell, 2014; Bryman, 1988; Hughes, 2012). In this research work, the inductive approach to quantitative study is been used and is related to the fact that the research starts with research questions, aims, and objectives that need to be achieved throughout the research (Dudovsky, 2011). Moreover, quantitative research is referred to empirical research studies, where researchers are more interested in understanding general principles of behaviours related to humans and animals, whilst attempting to estimate the average performance throughout the study. In addition, a thorough quantitative research protocol involving the collection of numerical data, usually on a smaller and more manageable scale (Blaxter, Hughes, and Tight, 1996; Bryman, 1997; McGregor and Murnane, 2010), which can then be generalised to a larger scale.

- **Reasons for Choosing a Quantitative Research Approach**

The quantitative research approach was selected for this study for a number of reasons, which are highlighted below. This approach provides researchers with more reliable measurements, and experimental control through sampling and design, with the possibility of generating causal statement and replicable processes. Furthermore, it allows the implementation of stronger analytical tests for validation purposes via statistical methods, implementing stratified cross-validation and logistic regression, and establishing feature correlations, deterministic and probabilistic predictions, etc., (Everitt and Hay, 1992; Hughes, 2012). However, Burns (2000) and Hughes (2012) questioned the fact that in the quantitative research approach, researchers are subjectively involved in the research process, which is regarded as a source of bias of the research results. This position is arguable in the sense that a researcher using a quantitative research will for sure collect data and process them using techniques such as AI techniques. The results always include the minimum intervention of the researcher. In addition, techniques are even employed to minimise the unwanted source of bias, including cross-validation techniques, balancing the dataset to avoid that the classifier identifies all the data points as belonging to the majority class, etc.

Nevertheless, this journey towards disease diagnosis has to be supported by a strong theory. The following sub-section proposes the Intelligent Technology Task Fit Model (ITTFM).

3.2.2. The Intelligent Technology Task Fit Model (ITTFM)

This sub-section outlines the development of the intelligent technology task fit model (ITTFM), which is an extension of the task technology fit theory (TTFT).

- **Intelligent Technology Task Fit Model (ITTFM)**

The ITTFM is a strategy developed to create an adequacy between the intelligent technology employed, and the task to be performed as proposed by Goodhue and Thomson (1995), a step

that is of paramount importance to the current research project outcomes. Hence, matching the SVA data visualisation performances and the dimensionality reduction, or the optimum support vector OSVM (optimisation of the support vector machine) classifier algorithm for improving biomarker discoveries in disease diagnosis. The ITTFM model can also be used to assess the potential of the multiple logistic regression (MLR) software predictive approach. This includes the probability to develop the NPC1 disease and the NPC liver dysfunction datasets, based on biomarkers detected.

The ITTFM model offers the possibility to use different experimental processes in order to assess and measure model performances using the looping processes, and a threshold value r_0. For example, NPC1 disease classification rates $\boldsymbol{r \geq r_0}$ could be used as valid assessment criteria. Where loop 1 is related to the OSVM classification, and loop 2 is related to the MLR probabilistic classification. The ITTFM model is presented below in Figure 1.

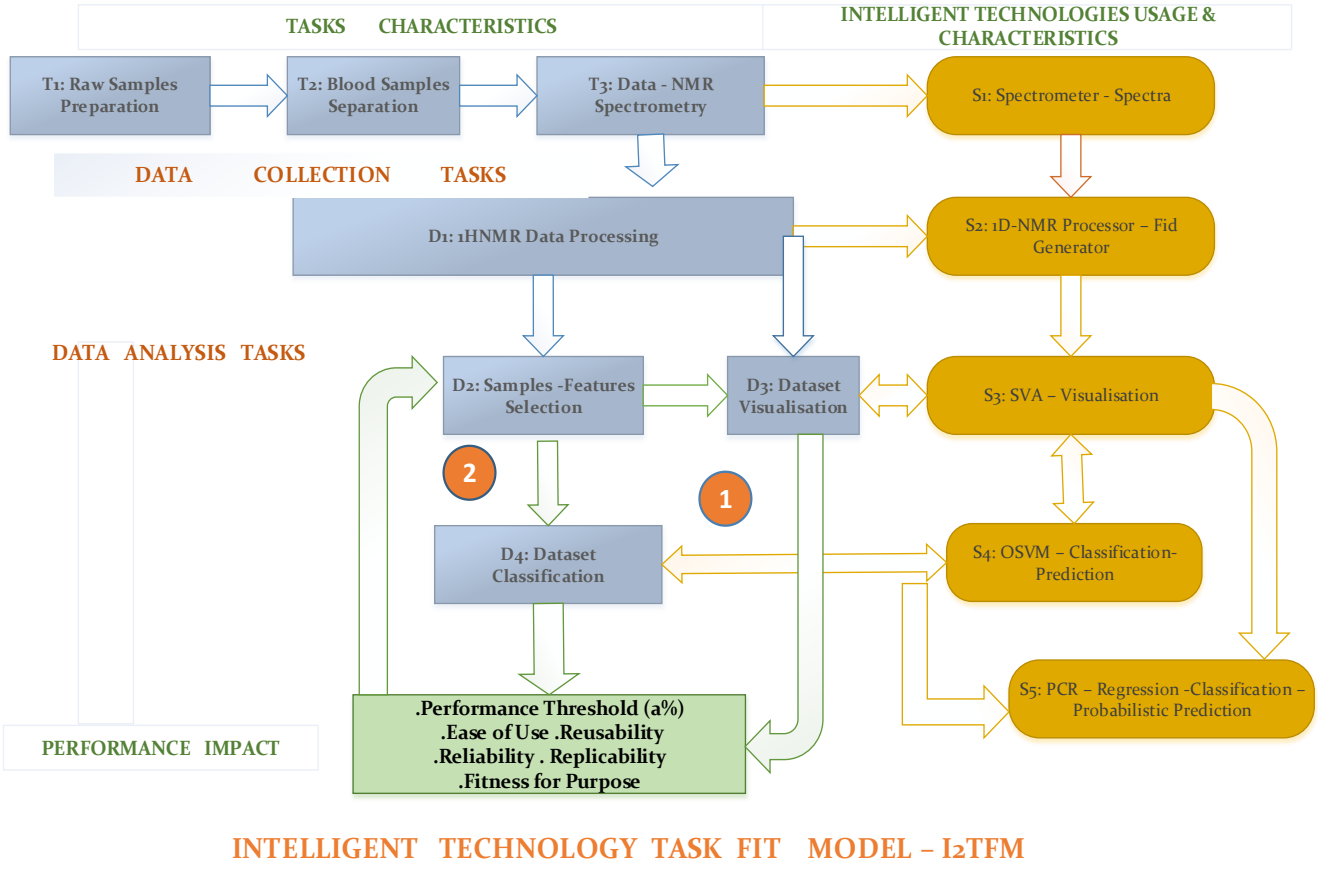

Figure 1. Different components of the Intelligent Technology Task Fit Model (ITTFM) with its three main components, including task characteristics, intelligent technologies usage and

characteristics, and the performance impact. The looping options to improve the system performance is defined by the stages 1 and 2.

- **ITTFM as an Extension of the TTFM**

ITTFM is an adopted model from the task-technology fit model (TTFM) developed by Goodhue and Thompson (1995). The model aims to apply intelligent technologies to IT/IS, for aligning intelligent technologies and the tasks to be performed intelligently. The intelligent technology is designed to fit the task performed, hence the descriptor intelligent technology task fit model. ITTFM supports the current research project development, and includes the following characteristics:

1. The looping processes 1 – 4 (as follows), allowing the researcher to go back as far as sample and feature selections involved in this research work, in order to perform an iterative and incremental assessment and gradually improving the model performance. Hence, when this criteria $\boldsymbol{r \geq r_0}$ is not met, the researcher can re-conduct the NMR data analysis process, change the sampling technique, or the features selection technique applied, etc. This means in clear that a trade-off has to be found between the quality of the tasks performed, and the intelligent technology model applied.

2. Implementation of a threshold r_0 allowing the researcher to validate the research overall process by comparing the performance rate r to r_0. For example, the threshold value depends on the type of model developed, and the kind of assessment carried out. Hence;

- A model developed to assess a **critical system** should have a higher threshold:

  $$r \geq r_0 = \boldsymbol{a} = 1 : thus\ the\ alignment\ should\ be\ perfect\ to\ handle\ the\ criticality$$

- A model developed to assess a **normal system** such as the classification of datasets to segregate between NPC1 disease patients and healthy control individuals can have a threshold set as follows:

  $$r \geq r_0 = \boldsymbol{a} = 0.8 : thus\ the\ alignment\ is\ set\ high$$

3. Self-assessment option for the final result improvement is another particularity of this model, which is implemented through the looping process combined with a threshold setting in order to incrementally improve the system performances in the form of a function

$r_n = r_0 + n\mu$ **(18)**; *where* r_0 *is the initial threshold,* μ *a constant which can be determined through experiment and n the number of loop applied.*

The table below gives a summary of two main types of systems and example of the threshold limits. This includes the model being used to assess a normal system for clustering, deterministic classification, probability and predictive classifications, etc. The same model with $r = r_0 = 1$ could be employed to assess that a patient's blood plasma sample is freed from early signs of the NPC1 disease, or more advancing secondary signs.

System Type	***Normal System***	***Critical System***
Threshold Limit r_0	$r \geq r_0$ (e.g. $r_0 = 0.8$)	$r = r_0 = 1$
Type of Usage	.Non-Critical .Assessing normal issues (Visualisation, Classification,..)	.Critical .System Assessing critical issues .High r_0 value for highest criticality

Table 1 Diverse types of system and the setting of the threshold value based on the type of model developed

4. Model flexibility: a certain level of flexibility is set for the model, with the number of loops and the threshold value, r_o, to be set, depending on the model developed, while μ will be a constant characteristic of the model developed.

5. The notion of alignment level and the level function l are introduced.
This refers to the correlation between a model's alignment to the task to be performed. Defining the correlation between the model performance and the level of its alignment to the task to be performed, this notion establishes that the higher the model performance, the higher its alignment level. Thus, a model achieving r = 0.75, for example, will be classified as level 3. The table below provides more details on the notion and level of alignment related to the ITTFM.

Alignment Level	***Level1***	***Level 2***	***Level 3***	***Level 4***	***Level 5***
Alignment Limits	$0 \leq a < 0.5$	$0.5 \leq a < 0.6$	$0.6 \leq a < 0.8$	$0.8 \leq a < 1$	$a = 1$
Alignment Type	Badly aligned 0% ≤ align ≤ 50%	Reasonably aligned 50% ≤ align ≤ 60%	Aligned 60% ≤ align < 80%	Strongly aligned 80% ≤ align < 100%	Perfectly aligned Align = 100%

Table 2. Relationship between the alignment level (x) and the model performance rate (a). The higher the value of the rate, r = a, the stronger the alignment.

To further consolidate this notion of alignment, a function has been defined that correlates the value of the performance level "a" to the alignment function.

- ✓ **The Alignment Function Model (AFM)**

$The\ alignment\ function\ \ l_x\ is\ defined\ as\ a\ function\ of\ the\ performance\ rate$ "a" $as\ follow$:

$\forall\ a\ \in [0,1],\ \ \exists\ x\ \in \{1,2,3,4,5\}\ \ f: a\ \rightarrow\ f(a) = x, and\ is\ noted:\ l_x = f(a) = x$

$which\ is\ read\ as\ follows$: $Given\ a\ performance\ rate, with\ a\ \in\ [0,1], there\ is\ a\ natural\ number\ x\ belonging\ to\ \{1,2,3,4,5\}; such\ that\ \ l_x\ is\ a\ function\ of\ the\ rate$ "a" $and\ is\ equal\ to\ x.\ Therefore, the\ following\ is\ verified$:

$a\ \in\ [0, 0.5[\ ,\quad l_1 = f\ (a) = 1 = level1 = Poor \longrightarrow Modle\ Badly\ aligned;$

$a\ \in\ [0.5, 0.60[\ , . l_2 = f\ (a) = 2 = level2 = Fair \longrightarrow Modle\ Fairly\ aligned;$

$a\ \in\ [0.60, 0.80[\ , l_3 = f\ (a) = 3 = level3 = Good \longrightarrow Modle\ Aligned\ ;$

$a\ \in\ [0.80, 1[\ ,\quad l_4 = f\ (a) = 4 = level4 = Very\ Good \longrightarrow Model\ Strongly\ aligned\ ;$

$a = 1,\quad l_5 = f\ (a) = 5 = level5 = Excellent \longrightarrow Model\ Perfectly\ aligned.$

$where\ 1, 2, 3, 4, and\ 5\ are\ values\ assigned\ to\ level1, level2, level3, level4\ and\ level5\ respectively, creating\ a\ relation\ between\ the\ level\ of\ alignment\ and\ the\ threshold$ "a".

Function 1. Alignment Function Model (AFM) creates a class function for the model alignment. This enable us to rank the alignment level from poor (level 1) - excellent (level 5).

- **The Notion of "intelligent - fit"**

The notion of *"intelligent - fit"* developed allowed the author to define other major criteria for judging the model's fitness for purpose, since it allows the model created to be fit for purpose and 'intelligence' simultaneously, and hence the reason for the name: *"intelligent - fit"*.

Meeting the model development criteria are conditions for it to be fit for purpose. In addition, the model should be intelligent, and this can be achieved by reducing the number of steps towards the final result, and most importantly improvements in its performance. This definition will be set as main criteria for judging the efficacy of the intelligent model developed. Therefore the notion of "*intelligent fit*" noted "*IF*" could therefore be presented as a function of the number of steps to get to the final result in modelling, the speed of the process and the alignment level between the technology and the task performed. Thus "*IF"* can be expressed:

$$\text{"f"} = \text{"}IF\text{"}\ , f = f(n, v, a);\ where, n\ is\ the\ number\ of\ steps\ towards\ the\ final\ result, v\ the\ process\ speed\ that\ is\ the\ number\ n\ of\ process\ per\ second\ t(s), and\ a\ the\ model\ performance\ rate.\ Therefore;\ given\ a\ model\ M_i\ member\ of\ class\ of\ Model\ M, the\ fitness\ for\ purpose\ can\ be\ written/expressed:$$

$$\forall\ M_i \subset M,\ M_i\ is\ \text{"f"} = \text{"}IF\text{"}\ \ /\ \ f = f(n, v, a)\ and\ M_i = M_{max}\ , where$$

$$M_i = M_{max}, if\ and\ only\ if\ n\ is\ minimum, v\ maximum, and\ a\ is\ maximum: c1.$$

$$v = \frac{n}{t} = v_{max}\ ,\ \ if:$$

$$1).\ n = increases, and\ t \to 0, or$$

$$2).\ n = increases, and\ t = constante \neq 0\ or,$$

$$3).\ n = constante, and\ t \to 0.$$

$$NB: Only\ the\ case\ 3)\ where\ n = cte = minimum, and\ t \to 0$$

$$respond\ to\ the\ condition\ c1\ defined\ above.$$

Function 2. The IF is defined as a function of the number (n) of steps toward the final result, the process speed (v) and the model performance rate (a), where the performance rate is equivalent to the level of alignment.

When a model meets IF conditions, the model developed fulfils the IF criteria; therefore, it is intelligent and 'fit for purpose'. In this case, the IF characteristics should facilitate model

validation. The number of loops and the value of the threshold to be introduced will depend on the project, which will ultimately enable determination of the quality of the IF.

Based on what precedes, a model will be defined as "intelligent-fit" if the number of steps taken to gain the best performance is reduced to the minimum, using the optimal algorithm. Therefore, the more intelligent the model, the higher the threshold, the faster the task is performed, and the better the alignment is. For this purpose, a process flowchart is presented below (Figure 2) explaining the different steps involved in using the ITTFM throughout this research, given that it can be generalised and applied elsewhere. The process flowchart below provides details regarding the overall modelling process, including the steps from the raw data collection to biomarker discovery, and finally pathway analysis.

- **Consideration**

n loopings noted L_1, L_2, L_3, L_4,…, and L_n show the potential of the modelling process to be designed to perfection, with the researcher having to re-start the process which is not meeting the validation criteria. Looping can re-start to higher level, i.e. L_2 can go back up to S_1 instead of S_4. In addition, thresholds could be set for several steps to be validated. This encompasses steps S_2, S_5, S_9, and S_{14}.

- **Intelligent Modelling Processes (IMP) for Biomarker Discovery and Validation**

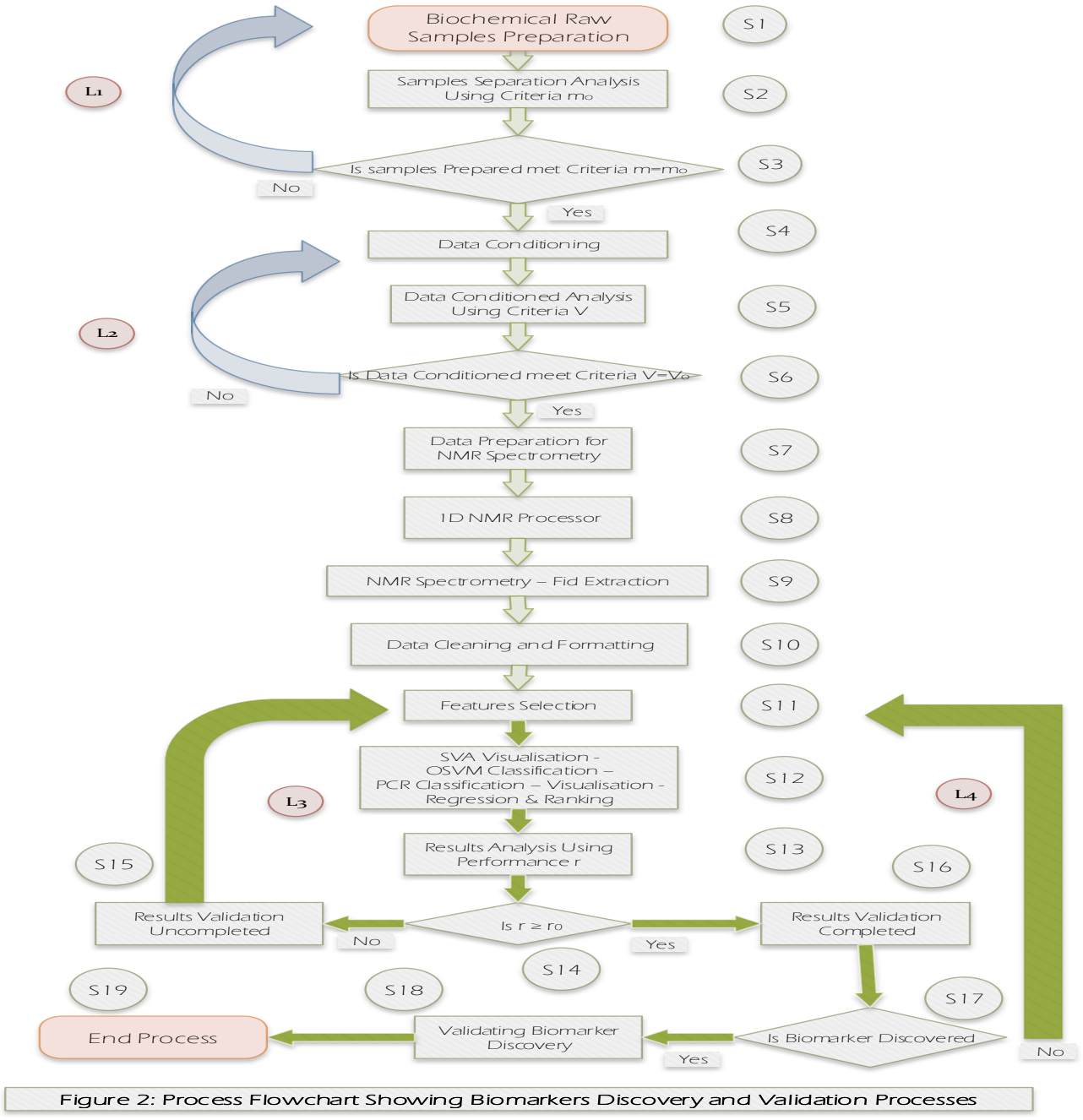

Figure 2: Process Flowchart Showing Biomarkers Discovery and Validation Processes

Figure 2. Process Flowchart and the step by step tasks for completion from the raw data collection, the biomarker discovery process and the pathway analysis. Overall, this shows how the IMP can be used to make the decision process more rational.

The process flowchart for biomarker discovery and the process sequence analysis presented above is said to be intelligent, because the model developed can re-assess itself and hence improves the performance value r, with r_0 being a threshold value in the development process. Therefore, if $r \geq r_0$, the result of the analysis is said to be valid. In addition, the "looping step"

includes some levels of flexibility, with the possibility of having more than one loop that is as much as necessary for the model to achieve maximal perfection. Different values (selected constants) have been employed in the sequence flow diagram and for this research, the constant m_o is the mass of the blood plasma samples; while constant V_o the volume of blood plasma utilised for ^{1}H NMR process: i.e. V_o = 5ml.

Assuming that the IMP model performance is set to r = r_0 = a = 0.8, i.e. level 4, and a very high alignment of the model with the task to complete, which involves biomarker discovery for the NPC1 and the NPC LDD datasets. Therefore, for a given model developed, acceptance could be judged by the author and or another research team using the IMP.

- **Consideration**

The validation stage could be combined with the ITTFM and its alignment model. However, this validation stage within the intelligent modelling process (IMP) is in line with the validation technique further discussed below.

Subsequently, an algorithm that allows the monitoring and validation of the different steps of biomarker discovery related to the IMP was developed. The example chosen here is related to blood plasma collected from NPC1 disease patients. Therefore, the IAD is used to validate the overall process involving data analysis and biomarker discovery. The algorithm even introduces the validation process of the biomarker(s) discovered and can be useful, ensuring that the entire process follows or is conducted according to standards. This process may give rise to improvements in the quality of biomarkers discovered. The different steps are marked i, where i is the step or the level. Hence, i = 1 is the starting level or 1st step corresponding to the raw sample preparation, i=2 the second step corresponding samples separation analysis, i = 3 the third step corresponding to the samples prepared and criteria m = m_0, etc. More steps can be added or removed from this process, depending on how the whole process is structured, i.e. the number of steps required for biomarker discovery and pathway analysis.

- **Intelligent Algorithm Design (IAD) for Biomarker Discovery and Validation**

The intelligent algorithm design is related to the steps involved in the independent statistical validation of the biomarkers discovered in this work. Researchers have demonstrated that predictive biomarkers for the possible response of patients to given treatments can be validated using extensive data, while prognostic biomarkers are related to existing diseases or to disease progression (Aronson, 2005; Buyse et al., 2010). They have also noted that whenever it is not possible to achieve statistical validation in view of requirements constraints, then surrogate end-points (i.e. substitute biomarkers) that call for a clinical validation may be applied. However, these researchers emphasised that there is no real consensus in the standards and processes to follow up regarding the statistical adoption of biomarkers (Buyse et al., 2010; Fan et al., 2013a; Rifai et al., 2006).

The technique proposed in this thesis is based on the process conducted during the experimental stages, and the results obtained during the analysis process. It should also be noted that the analysis stage 1 has included the scalar visualisation algorithm (SVA) and the support vector machine (SVM) techniques which could be performed separately. The same principle applies to the principal component regression (PCR) step. However, since the latter determinant position in the potential biomarker discovery, the PCR stage has been presented as a separate stage after performing visualisation and classification on the datasets. Depending on the manner in which researchers conduct their work, including independent statistical validation strategies, several options are available, and the algorithm presented below employs one of these differing possibilities.

Frame 1

1. Select NPC1 disease / NPC LDD patients
2. Collect blood samples / liver extract samples
3. Store blood / liver samples in vials for conditioning
4. Perform blood separation (plasma)/liver samples extraction
5. Measure volume v of plasma/liver extract
6. Is $v = v_0$?
7. $\begin{cases} -If\ Yes\ go\ to\ step\ 8 \\ -If\ No\ go\ to\ step\ 2 \end{cases}$
8. Apply NMR spectroscopy
9. Generate Spectra
10. Is spectra high quality $Q = Q_0$?
11. $\begin{cases} -If\ Yes\ go\ to\ step\ 12 \\ -If\ No\ go\ to\ step\ 8 \end{cases}$
12. Use 1D NMR processor and generate fid
13. Clean, format, and convert fid files to excel files
14. Apply features selection
15. Perform data analysis (1) = {SVA/OSVM}
16. Assess performance using the ratio r: Is $r \geq r_0 = a = 0.80$
17. Validate results: Are potential biomarkers visualised / level of classification result (excellent, good,..)?
18. $\begin{cases} -If\ Yes\ go\ to\ step\ 19 \\ -If\ No\ go\ to\ step\ 15 \end{cases}$
19. Perform data analysis (2) = {PCR}
20. Is potential Biomarker discovered?
21. $\begin{cases} -If\ Yes\ go\ to\ step\ 22 \\ -If\ No\ go\ to\ step\ 19 \end{cases}$
22. Administer miglustat as established drug for treating NPC1 disease.
23. Has concentration in biomarker1… biomarker (i) decreased?
24. $\begin{cases} -If\ Yes\ go\ to\ step\ 25 \\ -If\ No\ go\ to\ step\ 15/22 \end{cases}$
25. Validate biomarker1… biomarker (i)
26. End independent statistical validation process

Algorithm 1. Steps involved in the Intelligent Algorithm Development (IAD) for biomarker discovery and the independent validation process. The different steps involved can be modified depending on the way the research is conducted. In step 24, the researcher might decide to go back to step 15 to detect more potential biomarkers or just re-start the miglustat treatment/test in step 22.

Following this extensive development of the ITTFM, the next step deals with the research method used to support the present research project.

3.3. Research Methods

3.3.1. Notion of Research Methods

Research methods are different steps necessary to conduct the research process, whilst others consider them as a set of tools, techniques, or a component of research driven by procedures necessary for the collection and analysis of data (Clough and Nutbrown, 2012; Gabriel, 2011; McGregor and Murnane, 2010). In this thesis, the notion of research method is related to the techniques and procedures used in data collection especially data pre-processing and data analysis, which are in turn determined by the methodology applied (McGregor and Murnane, 2010; Wahyuni, 2012).

Two types of datasets were used during this process, and the data collection methods implemented involved in both cases a fieldwork, which was not performed by the author of this thesis, and a laboratory-based research work task. Raw data collected in the first fieldwork were those arising from blood plasma samples collected from NPC1 disease patients, and in the second case liver extract samples obtained from mice; these were in both cases further processed prior to data analysis. Therefore, data pre-processing related to ^{1}H NMR spectroscopy and 1D NMR processor was conducted. Moreover, noise, etc. was removed from such datasets. The features were labelled as NPC1 and WT (wild treated) for NPC1 and NPC LDD study.

For these reasons, the research methods that best suits this project is the quantitative research method. Several reasons have guided the choice for the research model, and they are been more detailed below.

3.3.2. Reasons for Choice of Experimental Methods

The experimental method was selected according to several rationales that are highlighted as follows. Firstly, the research project itself was a succession of empirical works. In reality, two parallel experimental processes were conducted throughout this research programme, and in terms of the logic and the rationale, they are consistent with the methods chosen. Data sources included the collection of data from NPC1 disease patients in the hospital in the National Institute of Health Clinical Centre in Bethesda, Maryland (USA). Secondly, datasets employed were all quantitative data, which were based on measurements. For example, raw data collected from NPC1 disease patients (blood plasma samples) or liver biopsy samples collected from experimental animals (mice) analysed via NMR spectroscopy and its associated 1D NMR processing software (McGregor and Murnane, 2010). Thirdly, the use of logical and more structured steps for data collection, which includes employing the software-based technique to generate finalised datasets (Bryan, 2004). In addition, data analysis proceed through structured and more controlled processes, and these included sample selection, feature selection, stratified cross-validation processes and the use of multiple logistic regression (MLR) as one of the data analysis techniques, etc.

Now that the term research method has been linked to the data collection process, an explanation on how the data collection steps were implemented is given. Moreover, sampling techniques are, of course, very important steps during data collection; these data were acquired by a collaborative research team, but for more clarity these steps are succinctly investigated in the next section.

3.4. Sampling Techniques

3.4.1. Introduction

A sample is a smaller but representative collection of units from an accessible population that is used to determine knowledge or truths about the population under study, and it should be free from any bias (Field, 2005; Kumar, 2011).

Equally important, are the sample selection techniques that assist the selection of a sufficient number of samples representative of the study population. Indeed, they allow the learning algorithm to become more accurate, reducing the search space, the time spent in training it, and finally the computational costs involved in the overall selection process (Wang et al., 2012).

There are different methods available to perform sample selection. Several options are presented below, while emphasising on the sample selection applied in this research project.

3.4.2. Sampling Techniques

There are two main sampling techniques, including probabilistic and the non-probabilistic samplings, whilst a possible third technique exists as a combination of these two approaches. Probabilistic sampling includes, amongst others, simple random sampling where each sample is given the equal opportunity of being chosen. This technique reduces bias related to sampling, ensuring that the sample selected is a true representation of the population of study (Blackstone, 2012). Another probabilistic sampling technique is the cluster sampling that randomly selects n clusters, and the samples within each cluster are randomly selected with equivalent probabilities (Blackstone, 2012; Grand Canyon University, 2015; Kumar, 2011; Sandelowski, 2000).

The non-probabilistic sampling techniques are mostly used for informal non-scientific research. They have not been used in this research, and as such will not be further investigated. Therefore, only probabilistic sampling techniques were used with the random clustering applied to biomedical datasets in order to reduce bias and make statistical inferences (Kumar, 2011; MEDICI, 2004; Sandelowski, 2000).

The correct implementation of the random sampling methods should allow researchers to make a sound prediction, to obtain reliable results that can be generalised with regards to the research findings. This, in turn, might pave the way towards successful data analysis, and bring the researcher close to the research objectives' achievement in respect to finding relevant biomarkers in the NPC1 disease diagnosis, since good quality data generates an effective data analysis (MEDICI, 2004; van Gulik, 2010). Below are provided more detailed justifications of the reasons that commanded the use of some of the techniques while choosing these samples.

3.4.3. Reasons for Choosing Sampling Techniques

Sample selection is extremely important in the data dimensionality reduction, in that it reduces the number of samples to be considered. Hence, reducing the processing, training time and the learning time for the algorithms utilised, and more importantly the cost of the overall process (Li et al., 2015).

Probabilistic sampling techniques, including random sampling and combination cluster random sampling were chosen to reduce bias related to sampling since each sample is given equal opportunity to be selected, i.e. equal probability of selection: EPS (Kumar, 2011). In addition, selecting samples in such a manner that they become truly representative of the population of study, positively influence the quality of metabolomics studies, helping in the generalisation of the results obtained, but more importantly drawing meaningful inference from the data analysis applied (Kumar, 2011; LASER, 2000; van Gulik, 2010; Vuckovic, 2012).

After applying relevant sample selection techniques in order to improve data analysis quality, the challenges faced during this research are now noted.

3.5. Research Study Challenges

This literature review shows that different studies were carried out for the purpose of biomarker discovery in diseases research in general, and also NPC1 disease and its more severe form related to liver dysfunction in particular. The lack of reliable biomarkers to trace the disease during the evolution phases remains a challenge (Fan et al., 2013). Furthermore, it is believed that more quality research needs to be conducted to address the lack of relevant biomarkers. In this study, challenges faced in research related to NPC1 disease diagnosis are highlighted.

3.5.1. Challenges of the Double Research Study

Different challenges were to be faced during this thesis. Among others were the following;

1). Typically, biomarkers discovery by the mean of computational intelligent technologies (CITs) in this Ph.D. project, is one of the main ways to inter-relate technology and medicine in disease diagnosis where biomarkers discovery is the main target. Moreover, the current study

focuses on the pathway follows by biomarkers for an insight into the diseases etiology, especially the NPC1 disease and the related liver dysfunction disease. Hence, CITs will be the tool helping for instance in measuring the probability of relating a given feature and the NPC1 disease progression (Fan et al., 2013).

2). Another challenge was to combine three main algorithms on the same objective and purpose of detecting biomarkers in disease diagnosis. The complexity being related to the ability of making sense of the different algorithms and techniques working together for a common goal.

3). Additionally, finding the same biomarkers or different biomarkers which can work together in understand the NPC1 disease etiology was another level of the challenge. Therefore, a lot of apprehension in the beginning of this research on how to make all this to work successfully was the main worry.

Nevertheless, what was considered at the start of this research as a challenge, quickly became one of the main reasons for implementing this current project. The possibility of finding biomarkers based on the diseases features to explain their progression and underlying causes make it very important to develop on top of the models, the features ranking strategy that was not in the agenda, i.e. the starting plan.

Finally, it was a real challenge at the beginning of this project to carry out research on disease diagnosis and especially, combining the NPC1 disease diagnosis and measuring the disease severity like the effect on the patient's liver. Although the two levels of research being dependent and strongly related to one another (same disease), adding to the fact that the same data modelling technique was applied to both datasets and having the biomarkers discovery as common objectives, the challenge remains because the research in disease was different from my original field of research in technology.

The data modelling technique used in this research also termed intelligent tri-modelling technique (ITMTs) encompasses the use of SVA for data visualisation; SVM for simple classification and prediction; and finally, PCR for classification and probabilistic prediction; while, biomarker discovery remains an aim on its own.

The Diagram below gives a general overview of the current and the future research, with the use of the same technological tools in the disease research; while biomarker discovery was bridging the gap between technology and medicine. The ITMTs allowing to detect different biomarkers, with biomarkers (1) for NPC1 disease diagnosis and biomarkers (2) for the diagnosis of the severe form of the disease related to liver damage. However, the future research will attempt to unify these biomarkers into biomarkers (3) = biomarkers (1) + biomarkers (2) discovery, that is aiming to find a unique biomarker noted biomarker (3) which will help to improve the NPC1 disease diagnosis, while detecting effective drugs.

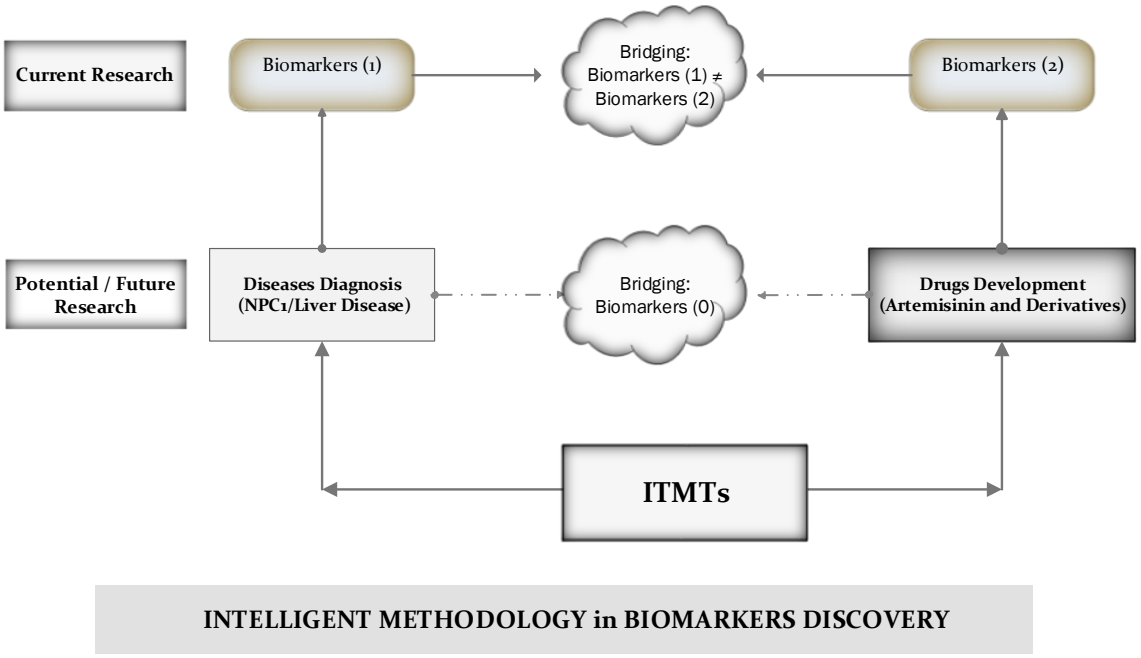

Figure 3. Intelligent methodology for biomarker discovery, bridging the biomedical and technological research projects. Research to be performed on the improvement of biomarkers discovery and drug-targeting is reserved for future studies. The intelligent methodology will help in bridging the gap between the disease diagnosis and the drug development in the NPC1 disease study.

3.5.2. Challenges in Biomarker Discovery for NPC1 Disease Diagnosis

Research in disease diagnosis, in general, and particularly the area of biomarker discovery, has brought a lot of excitement and enthusiasm within the research community. Many practitioners have noted an opportunity to tackle diseases more effectively and efficiently.

1) The ubiquitous computer and the related computerised systems have made it possible to generate a large amount of data as a result of metabolisms, which carry a lot of substantial information (Tzoulaki et al., 2014). Metabolomics and Metabolite profiling studies allow us to gain access to endogenous information in disease research. However, to date, the detection of biomarkers has not provided the "magic solution" for disease diagnosis which the research community expected. The reason being that biomarkers discovered could not reveal, uncover the "black-box" of occurrences at the molecular and endogenous level. This is valid for disease research, especially NPC1 disease. These have been linked to several explanations, including the poor level of information obtained from different biomarkers so far discovered (Fan et al., 2013).

2) Another level of challenge resides in the validation method of the research results, especially for biomarkers discovered in the NPC1 disease. For instance, it was noted that ***sphingolipid classes*** could be used as relevant biomarkers for NPC1 disease ((Fan et al., 2013). However, ensuring that new biomarkers carry more trusted information than the precedents so far discovered still represents a challenge.

3) A further challenge is related to the discovery of an adequate treatment to combat the NPC1 disease, as to date, only an alternative treatment is available. Indeed, miglustat (OGT 918, N-butyl-deoxynojirimycin) is a substitute treatment for the NPC1 disease and is approved in the UK and the USA. Typically, miglustat is directed to the inhibition of the synthesis of the glycosphingolipid, and as such can be used as a substitute to treat NPC1 disease patients (Fan et al., 2013).

4) The challenge attributable to the high expectation of finding appropriate treatments, is clearly related to the finding of appropriate, effective, and reliable new biomarkers for NPC1 disease treatment. The limited number of study subjects has affected the development of the method in the quest of new and appropriate treatments for those suffering from rare diseases, especially research based on clinical trials.

5) The reduced number of available patients to follow through the treatment process which can be prolonged, combined with the clinical outcomes' reliability, and its quantification which can be difficult. Indeed, it has not been possible to successfully classify all the NPC1 disease

patients, in view of different factors, such as symptoms difficult to identify, or patients unaware of their condition, etc., issues that need to be addressed (Fan et al., 2013). This has increased the pressure on researchers to come up with new techniques for monitoring NPC1 disease, including reliable quantification techniques for clinical trials focused on biomarker discovery.

Therefore, CITs' support in finding an improved response to NPC1 disease diagnosis is required, and this research attempts to provide an answer to this problem. Another level of challenge in this study was the possible outcomes of the research.

3.6. Possible Outcomes of the Study

Two main types of outcomes were possible for this research, this encompasses both expected and unexpected outcomes. More detailed explanation in this regard follows below.

3.6.1. Expected Outcomes

As expected outcomes, different possibilities were available for this research study and depending on the way the research was planned and conducted, the objectives were hence achieved. The expected outcomes are below listed:

1) The discovery of new biomarkers related to NPC1 diseases diagnosis;

2) The discovery of new biomarkers related to NPC liver dysfunction disease (NPC LDD), i.e. the severe form of the NPC1 disease;

Some of the unexpected outcomes of this project are also noted below.

3.6.2. Unexpected Outcomes

Different unexpected outcomes were produced during this research project. This includes,

1) The higher number of mathematical models developed to improve the current research process. This encompasses the Intelligent technology task fit model (ITTFM), and the Intelligent Modelling Processes (IMP) for biomarker discovery and validation.

2) The consequent high number of algorithms developed to support this research project. This includes the Alignment Function Model (AFM) creating a class function for model alignment, the Intelligent Algorithm Development (IAD) for biomarker discovery and the independent

validation process, and finally, the Intelligent Methodology for Biomarkers Discovery bridging the Biomedical and the Technological Research Projects.
Finally, after this presentation of the possible outcomes expected or unexpected in the present chapter, the next section gives a summary of this methodology chapter.

3.7. Chapter Summary

In this chapter, the chosen research approach, theory and methods are presented, and they supported the intelligent modelling processes (IMP) for biomarker discovery and validation. Thus, a quantitative research approach using empirical research method for such several reasons; including using quantitative data for research in biomedical, for research repeatability, reproducibility, replicability and reliability. Subsequently, the intelligent technology task fit model (ITTFM) was developed for matching the intelligent technologies to the tasks to be performed, such as improving biomarker discovery success. Moreover, several models and algorithms were developed to ease and structure the whole research methodology and research process. Furthermore, research challenges were highlighted, such as the challenge to validate the research methods and the results obtained. Finally, the outcomes of the study were succinctly noted.

"What you do makes a difference, and you have to decide what kind of difference you want to make."

Jane Goodall

4. RESEARCH DESIGN

4.1. Introduction

The previous chapter presented the methodology adopted throughout this research. Hence, the use of quantitative techniques as suitable research approaches was justified, followed by a well-detailed presentation of the intelligent technology task fit model (adopted from the Goodhue and Thomson (1995) task-technology fit model). This was completed by an explanation and justification of the research model adopted. In this manner, the experimental research method was confirmed as a suitable one for the present thesis. Moreover, the sampling techniques employed were visited to understand how achieving data dimensionality reduction could be decisive in data analysis, especially when a large amount of data becomes available thanks to high-throughput technology such as NMR spectroscopy. Furthermore, the different challenges faced during this study were highlighted, followed by the possible outcomes.

Vaus and Vaus (2001) considered research design as the step of structuring a research work in general, and particularly a scientific work undertaken by researchers, in order to help them to look for answers to research questions. They argued that research design is about the logical structure supporting the research study, helping researchers to plan and answer the research questions unambiguously. Therefore, it is necessary to identify the correct and suitable evidence that will support and shape the way to answer these questions. Hence, the way in which the research is structured and designed will depend on the fieldwork, and the empirical research planning and flow.

Research design is defined as a detailed outline of the way an investigation can be conducted. Typically, it includes data collection techniques, the necessary research tools employed, and finally the different means available for data analysis (Wyk, 2012).

The research design is the central component of the scientific inquiry, which allows the researcher to conduct a study in a way that reduces bias, distortion, random errors, error variance due to noise, and deals with the logical rather than the problem logistic (Nwankwo, 2004; Yin, 1989). Therefore, planning the set of logical procedures within scientific inquiry, are core, and help the researcher to obtain valuable data, or empirical evidence regarding isolated variables that are indispensable in conclusive inference about the constructs of the research questions. Consequently, certain researchers believe that the best research design is the one that best suits a particular problem at a given time (Heppner et al., 2007).

This research design chapter is organised as follows: Section 1) is an introduction to the chapter; Section 2) focuses on the mixed experimental research method that combines fieldwork and the lab-based research; Section 3) discusses data collection techniques and tools; Section 4) attends to data analysis techniques; Section 5) discusses the research validation processes; and finally, section 6) summarises of the research design chapter.

The next section to investigate is the mixed experimental research method, which is used to collect data required to address the aforementioned issues, including the pertinent research questions.

4.2. The “Mixed Experimental” Research Method

4.2.1. Introduction

The research method employed for this study corresponds to a “mixed experimental” research method. Several reasons support the use of this term. For example, a fieldwork was necessary to collect biofluid data samples in the United States of America. The blood sample collection from NPC1 disease patients was performed by a team of researchers in a hospital setting in

Maryland, USA. On the other hand, the NPC liver dysfunction disease (NPC LDD) dataset collection was performed at the University of Oxford, UK. For more clarification on the different steps of the mixed-experimental research method, the present sub-section encompasses an introduction and a more detailed analysis of the "mixed experimental" research method.

4.2.2. The "Mixed-Experimental" Research Method

The research method employed in the current Ph.D. thesis is termed as "mixed experimental" research method because it combines a fieldwork and a laboratory-based research method. This combined research approach is used to collect the entire data utilised in the current research project.

Regarding the biomedical data related to human blood plasma samples, they were collected from NPC1 disease patients and healthy control individuals under hospital settings, in the National Institute of Health Clinical Centre in Bethesda, Maryland (USA), by an American research team. The blood samples were then sent to Oxford University (UK), where it was finally turned into Excel files used in this thesis as part of the datasets involved in the current multivariate analysis.

The dataset associated to the NPC liver dysfunction disease (NPC LDD) was obtained from NPC liver dysfunction associated with this disease in a mouse model. The collection performed at Oxford University where mice were bred and housed under standard sterile conditions. A mutant mouse model of the NPC on a BALB/C was employed. The cohort encompasses NPC1 mutants, controls and the NPC1 heterozygous mice that were generated from the heterozygote mating. More detailed clarifications on the different stages mentioned above are given in the data collection session, while the rationale in using the mixed-experimental research method is highlighted below.

- **Rational of the "mixed-experimental" research methods**

The "mixed-experimental" research method was chosen for several reasons that are detailed as follow.

- **Based on fieldworks:**
- With regards to the first series of experimental research conducted, a team of researchers from the National Institute of Health Clinical Centre in Bethesda, Maryland, USA gathered the relevant data. This includes, the collection of blood samples from NPC1 disease patients and healthy control individuals.

- The second series of research-based biomedical study corresponds to the collection of set of data related to the NPC liver dysfunction disease (NPC LDD) samples. Mouse liver extracts were collected at Oxford University UK, by research group and following the protocol approved by the UK Animals Scientific Procedures Act 1986.

- The author of this thesis was not directly involved in these stages of the research related to the fieldwork. For this reason, no ethical issues were directly tackled in this thesis, especially those related to the blood samples collection from human participants and tissue collected from animals. However, these ethical requirements were confirmed by the research teams involved in the fieldwork.

- **Related to the laboratory-based research**
- The first laboratory-based research is related to the NPC1 disease study. The blood samples previously collected in the USA were further processed by team of researchers in the University of Oxford. The plasma samples separated from the serum in the blood specimen in a laboratory-based research, were used to produce NMR spectra using ^{1}H NMR spectroscopy technique, which in turn allowed to finally generate the biomedical datasets.

- The second laboratory-based research is related to sample collection from NPC liver dysfunction associated with this disease, which was performed on a cohort of NPC1 mutants, controls and the NPC1 heterozygous mice obtained from the heterozygote mating. A portion of the liver sample was weighed with an ice-cold extraction solvent that includes 1:1 of H_2O:CH_3CN with an equivalent of 20µL per mg of tissue. After homogenising mixture of solvent - tissue using an electric pestle rotor, and the homogenates obtained were centrifuged for 10 minutes at 4 ^{0}C at 10,000 x g. Hence, 1 mL of supernatant was lyophilised and reconstituted in 500 µl of the deutero-

chloroform (D_2O) containing 5.10^{-2} sodium 3-trimethylsilyl-(2,2,3,3-$^2H_4$)-1-propionate (TSP) and 50 µL of a 1 M solution of phosphate buffer (pH=7). Finally, before taken to NMR analysis, the samples were centrifuged and rota-mixed for 5 minutes at 5,000 x g at laboratory temperature, where the supernatants were transferred into NMR vials (5 mm) (Ruiz-Rodado et al., 2014).

- **Based on the use of metrics - Quantitative data**

- The laboratory-based research involved different processes such as blood plasma separation, NMR spectroscopy, in order to generate NMR data, define metrics, measurement, tests, etc., performed through stages that are normal steps involved in a typical experimental research strategy. Combined with relevant strategies for data collection, repeatable measurement steps were included in this research project (Bryan, 2004; Ross and Morrison, 1996).

- The final dataset used in this research are quantitative data for biomedical research. Through high-throughput technology such as NMR spectroscopy, large sets of data were generated in the laboratory-based experimental research with a systematic, repeatable and scientific approaches that help to build a reliable and more accurate model. These steps are in line with the experimental research method (Guo and Sheffield, 2006; Ross and Morrison, 1996).

- Converting the final NMR data (spectra) into numerical data was performed using software like the NMR spectrometer and the 1D NMR processor. The use of more rationale techniques that make the whole process repeatable, replicable and reliable. Indeed, employing techniques such as the intelligent binning, zero filing, baseline correction, etc…, allow a research team to reproduce this empirical research, adding to the fact that it becomes more reliable. These are criteria related to the logical and rational processes supporting the use of experimental research methods.

- **Based on the use of experimental research**

- Experimental research methods steps were employed to answer the research question effectively and to meet the research aims and objectives. This includes the discovery of biomarkers.

- However, the use of experimental research methods allows the author to discover the major biomarkers for NPC1 disease diagnosis while establishing correlations between disease features and the disease presence (Brown, 2010; Liang and Fang, 2006).

- One should note that in the particular case of this research, different combinations were possible during the experiments, giving that different models were compared against each other. For instance, with regards to SVM and MLR, the dataset standardisation application to the model developed was an option available. In addition, to build the model, one of these two kernel functions could be used, i.e. the linear or radial basis function (RBF) kernel. Moreover, the feature selections provide the possibility to choose between different options to build different models, and their performances were compared against each other. This includes the nearest neighbour for feature selection and the option of varying the number of features used. Furthermore, the possibility of oversampling the original dataset to correct the issues related to an imbalanced dataset was performed by using the ADASYN method that facilitates the learning process in the case of imbalanced dataset (He et al., 2008).
  The model improved the detection of biomarkers in both studies that was clearly identified and presented.

In conclusion, this research which combines two different types of experimental research methods, which are the fieldwork and the laboratory-based research is naturally and logically more suited to be term as "mixed-experimental" research methods.

4.3. Metabolomics Data Collection Techniques

4.3.1. Introduction

During data collection, the techniques employed will help to determine the quality of data that are at a researcher's disposal. Therefore, they will help to pave the way to address the research

problem (Priyan, 2012). Data collection techniques are key steps in the structure of a research study because the result of the research depends on its suitability, which could influence the validity of the research undertaken (Pratiwi, 2013).

4.3.2. Data Collection Tools

Different tools, especially intelligent algorithms were utilised during data collection methods. However, it should be noted that the tools exposed in this sub-section including the NMR spectroscopy and the 1D NMR processor were the main tools involved in the collection of the biofluid metabolites.

- **NMR Spectroscopy as Metabolomics Data Collection Tool**

The NMR phenomenon was discovered independently, but concurrently by different groups of researchers in 1946. It is related to the study and the application of certain properties of atomic nuclei; among which the energy variation, especially when the nuclei is placed in a strong external magnetic field, and subjected to Radio Frequency (RF) irradiation. Thus, different nuclei will absorb RF energy and resonate at different frequencies in the view of the electronic environment surrounding them. This resonance frequency difference compared to the standard in the magnetic field, known as chemical shift, implies a different level of energy absorption and energy emission (De Graaf, 2007).

NMR spectroscopy is a non-invasive, non-destructive and rapid technique used to study a plethora of metabolites contained in various metabolomics samples such as biofluid datasets (Siddiqui, 2003). NMR spectroscopy is based on the fact that metabolites nuclei exposed to an intense external electromagnetic field $\overrightarrow{B_0}$, align their magnetic moment against the field $\overrightarrow{B_0}$ to gain a higher energy state or with the field $\overrightarrow{B_0}$ to attain a lower one that is overall preponderant (De Graaf, 2007; Siddiqui, 2003). However, atomic nuclei absorb energy at different resonance frequencies due to the energy variation within their vicinity; therefore nuclei in a lower energy state may go to a higher energy state, orienting their magnetic moment against the field $\overrightarrow{B_0}$ when RF energy is brought within the nuclei's magnetic micro-environment.

Consequently, the majority of the ^{1}H nuclei from different clusters samples will go to their excitation state corresponding to their higher energy state by absorbing RF energy in the region of the spectral frequency interval of [0-10] ppm (parts per million) (Siddiqui, 2003). The absorption is followed by energy emission (return to a stable state) in the form of photons. The energy E absorbed by the ^{1}H nuclei is proportional to the RF, and can be expressed as follows:

$$E = h\,\nu \quad (\mathbf{18}), with\ h\ corresponding\ to\ the\ Planck\ constant\ and\ \nu\ the\ RF.$$

$$This\ means\ that\ E\ is\ a\ function\ of\ the\ frequency, \qquad that\ is \quad E = f(\nu). \quad (\mathbf{19})$$

The NMR data so generated are presented as a spectrum, which is a signal representing the energy amplitude against the frequency; with the ^{1}H spectrum located in lower and higher frequency region 0-10 ppm (De Graaf, 2007; Siddiqui, 2003). A signal is usually referenced from an organic solvent such as the tetramethylsilane (TMS) that is used as a chemical shift reference (chemical shift of TMS = 0.00 ppm), or a water-soluble derivative TSP, which also appears at 0.00 ppm (Siddiqui, 2003).

At resonance, the intensity of the signal obtained is proportional to the product of the number of magnetically-equivalent nuclei present in the functional group, together with the concentration of the molecules present in the group (Siddiqui, 2003). The NMR spectra produced will now be passed on to the 1D NMR processor to be further processed.

- **The 1D NMR Processor**

The 1D NMR processor is a software with different features and functionalities, making it a core tool for the current study, especially for laboratory-based research. In fact, the 1D NMR processor was used to perform several transformations that improve the overall quality of NMR spectra. Amongst these, the Fourier transformation, zero filling, phasing, baseline correction, spectrum editor and covariance NMR, peak picking, etc., (Softpedia, 2014). Some of these keys features are highlighted here. The baseline correction corresponds to the step where the distortions caused by the hardware artifact like the NMR spectrometer, and / or due to the

solvent's high concentration were removed. However, phasing allows minimising the discrepancy between resonances and their alignment, especially those from zero-order and first-order due to several reasons such as the time-delay between the nuclei excitation and the response reception (Ravanbakhsh et al., 2015). Furthermore, the zero filling technique that consists on appending zero points to the free induction decay (fid), is performed before a Fourier transformation to increase the digital resolution of the spectrum (de Graaf, 2007; Keeler, 2004). Through the Fourier transform, the time-dependent free induction decay (fid) is converted into a frequency dependent spectrum. Finally, 1D NMR can perform a structural elucidation of protons close or adjacent to different functional groups present in the spectral data samples.

- **Consideration**

These techniques are utilised to improve the quality of the spectra produced (see section 2.4 in Chapter 2). Indeed, the applied chemistry development (ACD) 1D NMR processing software allows further processing of the fid file into excel files ready for analysis. In addition, the ACD 1D NMR has been used to give a clear picture of compounds complexity; this includes understanding the bonding between atoms inside the molecules while suggesting possible chemical structures.

Spectrum quality is affected by the presence of noise. This, in turn, can affect and be detrimental to the metabolomics profiling of small weighted biomolecular concentration (Xi and Rocke, 2008). To tackle noise in the spectrum, different means such as the polynomial fit and spectrum averaging methods are applied. More detail on these techniques can be found in (Halouska and Powers, 2006). Generally, noisy regions in the spectra are removed via intelligent bucketing that allows automatic splitting of the spectral regions in the interval, in such a way that a spectrum is not divided into adjacent intervals (Halouska and Powers, 2006; Worley and Powers, 2015).

- **Consideration**

Other tools were involved in this research, especially those related to data analysis, and these have been included in the data analysis section for more clarity.

4.3.3. Data Collection Techniques

- **Introduction**

The data collection techniques applied in this research work are typically related to the handling of the biomedical data samples throughout the fieldwork, and the laboratory-based research. However, because the biomedical data were provided by collaborative research teams, a brief reference will be made to this data collection process. Finally, this sub-section includes the following: 1) an introduction; 2) biomedical data collection related to NPC1 disease; 3) biomedical data collection related to NPC liver dysfunction disease (NPC LDD); and 4) molecular assignment of metabolites.

- **Biomedical Data Collection Related to NPC1 Disease**

This data collection involved gathering data from NPC1 patients and healthy control individuals, which were further processed via NMR spectrometry and 1D NMR processor to generate metabolomics dataset (see sub-section 4.2.2).

- **Biomedical Data Collection Related to NPC Liver Dysfunction Disease (NPC LDD)**

The data related to the NPC liver dysfunction associated with this disease is obtained from mice and performed at the University of Oxford where the mice were bred and housed under sterile conditions. However, the ^{1}H NMR spectral acquisition took place at De Montfort University, School of Pharmacy. The spectra were generated using a Bruker Advance AM-400 spectrometer, which operated at 399.94 MHz, 298 K as probe temperature. The spectral acquisition used TOPSPIN version 3.0, Bruker Biospin, with a relaxation time of 3 s, with a spectral width of 4,800 Hz. The advanced chemistry development ACD software was employed for baseline correction, zero filling, and the binning technique applied. The intelligent bucketing was applied to avoid splitting the same signal between two bins or buckets. The technique used an algorithm to calculate and determine the maximum size for the spectra obtained, which are between 0.02 to 0.06 ppm (Ruiz-Rodado et al., 2014; Sousa et al., 2013; Vaidyanathan et al., 2006).

- **Molecular Assignment of Metabolites**

For both type of datasets used in this study, the molecular assignment of the ^{1}H NMR spectra collected was performed using reference to appropriate materials and publications, and it was double checked with the human metabolome database. The spectra identified were generated following the process mentioned above.

Now that the NPC1 and the NPC LDD data collection were succinctly mentioned, the different stages of the data multivariate analysis are detailed below. The Intelligent Tri-Modelling Techniques (ITMTs) were implemented for the data analysis purpose, meet the research objectives and answer the research questions. The ITMTs development has for objective the improvement and optimisation of both processes related to the discovery of biomarkers in the NPC1 and the NPC LDD diagnosis.

4.4. Intelligent Tri-Modelling Techniques for Multivariate Data Analysis

4.4.1. Introduction

In this thesis, the Intelligent Tri-Modelling Techniques (ITMTs) are developed and applied in the multivariate data analysis section. No research study to the author's knowledge has so far used these Intelligent Tri-Modelling Techniques to analyse datasets in biomedical research area. Developed around three main algorithms in machine learning and data classification, the model detailed below is typically based on data visualisation with SVA, classification with SVM, dimensionality reduction and predictive analysis with PCR. In this research, PCR has been presented as a combination of PCA and MLR in data analysis. The ITMTs are involved in the NPC1 disease diagnosis, conducted in parallel to NPC1 liver dysfunction disease diagnosis. Therefore, the ITMTs combine the advantages of the three models taken separately, as noted above. Additionally, the intelligent tri-modelling technique takes into account the interaction and relationship existing between variables or candidate biomarkers inherent to a multivariate study (Marcello Manfredi, 2013). This section therefore includes the following:

1) an introduction; 2) the opportunities offered by SVA and the PCA to visualise data and determine whether they separate effectively between different classes, before they can be further investigated for modelling. In 3), the possibility to perform a deterministic classification using SVM with regard to biomedical datasets. Finally, in 4) the potential presented by the PCR algorithm using PCA to determine the principal components (PCs) and reduce the data dimensionality. This is followed by the MLR for probabilistic prediction in order to establish a strong correlation between features, and the risk of developing at an early stage the NPC1 and its associated liver dysfunction disease (Taooka et al., 2014; Wijburg et al., 2012; World Wide Antimalarial Resistance Network, 2012). The use of PCA to determine the principal components or factors and carry out the analysis based on these uncorrelated principal factors can lead to the selection of the potential biomarkers. Indeed, the factor loadings and scores can help respectively to determine exactly the correlation of the original features to the principal components, and conclude on the contribution of every feature in diseases diagnosis and perhaps prognosis (Marcello Manfredi, 2013).

4.4.2. Scalar Visualisation Algorithm (SVA) for Data Analysis

- **Theory**

The term visualisation is related to the creation of images that carries information about the relationship existing within data in order to gain insight into the underlying data structure (Johnsona, 2012). There are different types of visualisation algorithm, including the scalar algorithm, the vector algorithm, and the tensor algorithm (Komura, 2016). The scalar visualisation algorithm (SVA) that is the focus of this sub-section works by converting scalar data type into coloured data. Thus, the ^{1}H NMR resonances produced from the NPC1 and NPC LDD metabolomic data is turned into an index that is, in turn, related or assigned to a colour selected from a colour lookup table. In this manner, every single signal value (scalar data) is mapped to a particular colour in the lookup table, in such a manner that underlying transformations at the molecular level can be displayed and visualised. Therefore, the scalar visualisation algorithm (SVA) can change the data type, for example, changing scalar data into a colour data type (Johnsona, 2012). The same researcher believes that visualisation is a technique that can help us to understand large and complex information carried by data in

general. For this reason, the technique has been applied in different research domains, including computer science, engineering, biology, particle physics, etc., (Johnsona, 2012).

- **SVA Model Applications**

The scalar visualisation algorithm has been applied in different fields of research and some of these applications are presented here. In the following study, scalar visualisation algorithm was applied in the surgical field where surgeons experienced visualisation of the cerebral vessel structure in order to have a clear view of disease diagnosis (Luo, 2013). To overcome the problems related to the traditional ray casting of the three dimensional view, which doesn't allow surgeons to establish a clear distinction of the vessels depth, this research proposes a distance colour blending and the stereoscopic depth enhancement that allows us to improve the depth perception of vessels that have complex structures (Luo, 2013).

Another application of the SVA is related to the adaptive surface visualisation based on the contour measurement technique for the assessment of the cardiovascular disease (CVD). The approach developed based on different visualisation techniques could improve the vessel surface shape and depth visualisation. It is important in the CVD diagnosis process to perform a clear visualisation exploration of the blood flow information in the vessel surface area in order to understand the blood flow behaviour closely related to the CVD. Problems such as vessel occlusion are common in the case of CVD, and hence blood flow expert visual expertise is required. The information required by these experts include quantitative and qualitative information that can help to overcome the complexity of the diagnosis, in turn related to the complex natures of data to be analysed (Lawonn et al., 2014).

- **Presentation of the SVA Algorithm**

Frame 2

1. *Select Dataset = {**Plasma**, Liver}*
2. *Select Dataset Normalisation Option = {**ON**, OFF}*
3. *Select Modelling Type = {**SVA**, etc.}*
4. *Range Scalar Value = Range Colour*
5. *Assignment:*
   - *Assign Peak minimum value = Blue (Colour)*
   - *Assign Peak maximum value = Red (Colour)*
6. *Colour Look-up Table : Scalar value convert = Index*
   - *Scalar Minimum = Peak Minimum = Minimum Index = I_{min}*
   - *Scalar Maximum = Peak Maximum = Maximum Index = I_{max}*
   - *Scalar Value = Peak Value = Index Value = I_i*
   - *Scalar Value Range = {I_{min},……, I_i, …., I_{max}}*
7. *Colour Range ={colour0, colour1, …., colour n-1)*
8. *Colour Mapping*
   - *If Ii < I_{min} , colour = $colour_{min}$ = colour0 = C_0*
   - *If Ii < I_{max} , colour = $colour_{max}$ = colour $_{n-1}$ = C_{n-1}*
   - *Else use colour transfer function $f_t(I_i) = C_i$*
   - *NB: Combine colour red, green, and blue to generate different colour based on peak intensity = index value I_i*
   - *End If*
   - *Assign Index to Colour*
   - ***End Colour Mapping*** (Komura, 2016)

Algorithm 2. SVA algorithm for data visualisation for the intelligent tri-modelling technique (ITMTs), with possible steps following in colour mapping.

4.4.3. Support Vector Machine (SVM) for Data Analysis

- **Theory**

The classification tasks involve, amongst others the necessary task of dividing the original dataset into training and testing sets. A kernel function is used to map the training set onto higher dimensional space, where a hyperplane can be found to separate the sample into two or more distinct classes (McDermott et al., 2013).

In this manner, the model is trained to segregate between two classes C1 (+1) and C2 (-1). For example, segregating between diseased and healthy control individuals in the NPC1 disease dataset classification. After the training step, the model is ready to classify unknown data between the two classes C1 and C2, with each class containing a target value of +1 or -1. The SVM has to produce a model that can ultimately predict the target value of every single instance (sample) in the testing dataset based on its features (Marcello Manfredi, 2013; Michie et al., 1994). When such a hyperplane exists that separates the two classes C1 and C2, therefore, a vector w and a scalar b verifying the inequalities exist and is expressed by:

$Given\ the\ labeled\ training\ set\ (x_i\ , y_i), \dots\dots\dots (x_l\ , y_l),\quad y_i\ \{-1, 1\}\ :(\mathbf{20})$

$w.x_i + b\ \geq 1 \qquad if\ y_i = 1$

$w.x_i + b\ \leq -1 \qquad if\ y_i = -1\quad :(\mathbf{21})$ (Cortes and Vapnik, 1995).

In the case where the two classes are linearly non-separable, then the problem requires the finding of a hyperplane that will carry out the separation with minimum error. However, the problem at hand becomes an optimisation problem presented as follows, considering the following pairs of instance:

$(\ x_i, y_i),\ i = 1, \dots\dots, l, with\ x_i \in \mathcal{R}^n\ and\ y\ \in \{1, -1\}^l : (\mathbf{22}).$

The SVM algorithm has to find a solution to the given optimisation problem

$$\min_{w,b,\xi}\quad \frac{1}{2}w^T w + CF\sum_{i=1}^{l}\xi_i\ :(\mathbf{23});\ where\ C\ is\ a\ constant, F(u)\ with\ F(0) = 0\ is\ a$$

$monotonic\ convex\ function, and\ \xi_i is\ the\ sum\ of\ deviation\ of\ training\ errors$

$Given\ that\ \ y_i(w^T\phi(\ x_i) + b\ ) \geq 1 - \xi_i,\quad with\ \xi_i, \geq 0.$ (**24**) (Cortes and Vapnik, 1995).

- **SVM Model Applications**

The support vector machine is a deterministic classification algorithm using a heuristic method for prediction. SVM is based on the principle of detecting a hyperplane that is used as a separation n-dimensional plane between two different classes (Verplancke et al., 2008). The process of defining a hyperplane can be achieved through a straightforward process or through a transformation using a linear kernel function, especially when it is not possible to find such hyperplane in a lower dimensional input space, by mapping the dataset into a higher dimensional features space. Such a transformation can allow the detection of the separation hyperplane (Grootveld, 2014; Salazar et al., 2012; Zhang et al., 2006).

In the specific case of this binary classification, SVM has been employed in order to discriminate between diseased subjects and healthy control ones in relation to two biomedical datasets, the NPC1 diseases dataset, and liver dysfunction disease dataset.

- In this project, the SVM uses different setting with the objectives of optimising the classifier performances. In this regard, the LibSVM 3.21 folder is added to the search path to avoid using the default SVM function provided by Matlab. This choice allows the setting that implements the SVM optimisation and it encompasses the following options: the dataset standardisation, type of kernel applied, i.e. linear or radial based function (rbf), repeated k-fold stratified cross-validation, etc. The main purpose of which is to have data of the same scale in order to proceed to an effective comparison between them. The result of this optional setting, the auto-scaling will generate a dataset with zero mean and a unit variance (Jajuga and Walesiak, 2000; Raschka, 2014).

- The choice was given for the choice of kernel functions used, including the linear or the RBF kernel. However, the method used in this research to estimate the classifier performances, was conducted through the repeated k-fold stratified cross-validation. The k-fold Stratified Cross-Validation provides more reliable estimates than when using the single cross-validation technique. Stratification handles the bias and variance more successfully than the single cross-validation method. Furthermore, repeated

cross-validation compared to cross-validation allows us to attain a more precise estimate of predictive performance (Hastie et al., 2008; Kuhn and Johnson, 2013). Hence, estimation of the accuracy without confidence interval has no real value. The confidence interval is defined to give more credit to the values of the classifier performance (Kohavi, 1995). Different values of k can be chosen, for example k Є {5, 10}, with 10-fold stratified cross-validation being more effective in terms of model selected for real world datasets (Kohavi, 1995).

- Samples selection applied is based on the ADASYN technique to implement the oversampling techniques, in such a manner that synthetic samples in the minority class are created to balance the dataset.

- With regard to features selection, a certain number of nearest neighbour parameters are used to perform the ReliefF, a features selection method that improves classifier performance by removing redundant features (Rosario and Thangadurai, 2015; Zhang et al., 2016). Hence, it is a feature elimination technique to remove the less significant features in this binary classification process (Guyon et al., 2002), combined with the repeated k folds stratified cross-validation technique, where the dataset is divided into a total of 5-10 folds. During the modelling process, (k-1) is used for training and 1 (one) fold is used for testing the model, and this process is then repeated n times, so that every single set is used at least once as training and testing sets. The technique is used to validate the whole modelling process (Grootveld, 2014). The algorithm used to develop the process described is now outlined below:

➢ **Pseudo-code: Adapted SVM**

Frame 3

1. *Including LibSVM*
2. *Select dataset = {Plasma , Liver}*
3. *Select data standardisation option ={ON , OFF}*
4. *Apply grid search technique for best combination (C , γ) and optimum OSVM*
5. *Select k folds = {5 , 10}*
6. *Apply k folds cross-validation technique*
7. *Select classification function = { Linear, RBF}*
8. *Select number of nearest neighbour for ADASYN method = {10 , 15 , 20}*
9. *Candidate SV = {closest pair (x_i , x_i'} from opposite classes}*
10. *Select number of nearest neighbour for features selection = {10 , 20}*
11. ***While** there violation points **do***
12. *Find violator*
13. *Repeat = {10 , 20}*
14. *Candidate support vector SV = Candidate SV U violator*
15. ***If** any $\alpha_p < 0$ due to addition of c (another SV) to S (support vector set) then*
16. ***N.B.** $f(x) = \sum \alpha_i . y_i . K(x_i, x) + b$, where f kernel associated with the SV and $x_i \in S$*
17. *Candidate SV = Candidate SV\p*
18. ***Repeat until** all such points are pruned*
19. ***End If***
20. ***End While*** (Akay, 2009; Blagus and Lusa, 2012**;** Vishwanathan and Murty, 2002)

Algorithm 3. SVM model combining grid search technique, cross-validation, and features selection techniques for the intelligent tri-modelling techniques (ITMTs).

- **Consideration**

In the present thesis, principal component analysis (PCA) is combined to multiple logistic regression (MLR) to generate the principal component regression (PCR) model. These algorithms are presented in the very order as previously mentioned, i.e. PCA first and then PCR, and they are next to follow in this thesis.

4.4.4. Principal Components Analysis (PCA)

- **Theory**

The Principal Component Analysis (PCA) model is a descriptive method that presents the original variables in a new reference system characterised by principal components. PCA involves the projection of the original features onto orthogonal planes in order to generate new features or factors, reducing the variables space. Therefore, PCA is as well-known data dimensionality reduction technique (Marcello Manfredi, 2013). Hence, PCA involves a change in variable space, transforming variable space into factor space. In the samples' space, the vectors length correspond to the variability of the variables, while the angle between vectors represents the correlation between variables, and is usually expressed as the correlation coefficient α for standardised variables (Amath 301, 2016; XLSTAT, 2010). The factors (components), also called eigenvectors, are orthogonal and non-correlated, and allow an explanation of the variability in the original dataset based on the main PCs (Kumar and Kalra, 2016; Mahmoodabadi et al., 2008; Nikas and Low, 2011; Rustempasic and Can, 2013).

However, PCA allows visualisation of the correlations between different features. In fact, PCA allows the plotting of the data on the principal components (F1, F2, F3 ...) in order to visualise whether the different classes C1 and C2 separate well. In this case, further study can determine the main features responsible for this total separation given that the principal components are linear combinations of the original variables (Marcello Manfredi, 2013; McDermott et al., 2013). This may provide valuable information regarding the possible presence of biomarker(s), which might lead to a possible understanding of the underlying transformation at the molecular level with respect to metabolites involved in disease process.

The major reason for the application of PCA is related to the selection of the principal components (PCs). More detail explanations are provided below.

- **PCA Model Applications**

Different cases have been reported in the literature in relation to the application of PCA as a visualisation tools to find out whether there is total separation between classes in the PCs' space. With respect to the understanding of the likelihood of a patient developing diseases such as the NPC1, PCA can provide very useful information regarding candidate biomarker

detection (Marcello Manfredi, 2013). PCA was also employed for the identification of PCs that could facilitate the identification of major biomarkers for disease diagnosis and prognosis (Cluzeau et al., 2012; Wang, 2008).

In the present study, PCA is used to define the main factors that are the axes upon which the dependent variable Y will be regressed. In this manner, the factors or PCs are expressed as linear combinations of original disease features. This has the potential to reduce the disease feature space from 55 original features to 10 main ones for plasma datasets, and respectively from 143 to 10 for liver dysfunction dataset. This dataset dimensionality reduction has the potential to ease and improve model performances. With regard to the selection of the variables involved in biomarker discovery, the computational time, and the costs involved can be also reduced (Scannell et al., 2012). In the case of this study, the PCs selected explain more than 90% of the total variability in the original dataset (Wang and Abbott, 2008; Gauderman et al., 2007). Finally, the PCs selection was performed using XLSTAT 2016 software, and the following pseudo-code highlights the different steps followed.

- **Pseudo-code for PCA in XLSTAT 2016**

Frame 4

1. *Select dataset = {Plasma, Liver}*
2. *Select PCA type = Pearson (n)*
3. *Option = No Filter & No Rotation*
4. *Supplementary data = {No observation}*
5. *Supplementary variables = {No supplement variable}*
6. *Data option = {No Missing data}*
7. *Input = {Descriptive statistic and correlation}*
8. *Output = {Eigenvalue, Factor loading, variables/factors correlation}*
9. *Factor scores, contributions, squared cosines*
10. *Charts = {Observation, Labels and coloured labels}*
11. *Select Ok*

End PCA

Algorithm 4. PCA algorithm implementing principal component selection for the intelligent tri-modelling techniques (ITMTs). This allows us to understand the use of PCA as an aid to diagnosis. The Pseudo-code exposed is related to the XLSTAT 2016 version software used in this research.

- **Consideration:**

Based principally on the factor loadings and the factor scores, the principal components (PCs) selected undergo principal component regression (PCR) in order to determine the potential biomarkers. Further studies may be required to precisely determine the major potential biomarkers. Nevertheless, the MLR that is next visited is required to complete this process.

4.4.5. Multiple Logistic Regression (MLR) for Data Analysis

- **Theory**

Multiple logistic regression (MLR) is a probabilistic classification algorithm using a binary nominal dependent variable Y to predict the outcome of an event such as disease likelihood of appearance or progression, etc., based on the independent variables X_i, which can be correlated or uncorrelated. In this regard, multiple logistic regression (MLR) is a statistics model used in data analysis, with one binary nominal variable (dependent variable). For example, the value (+1) can be attributed to diseased, male, yes, etc., and the value (0) for healthy, female, no, etc., and two or more measurements which are the independent (Schoonjans, 2016). The variation in the measurements X_i (which can be correlated) permits prediction of the probability of the dependent variable Y. In this manner, MLR determines/defines the best fitting model describing the relationship between the dichotomous characteristic of interest (dependent nominal variable = outcome variable) and a set of independent or non-independent (explanatory) variables. The probability to predict an outcome is provided by the equation (**25)**:

$$Y = Y_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 \ldots + \alpha_n X_n : (\mathbf{25}). Higgins, (2005) shows\ that\ the\ Y_0$$

$$intercept\ of\ Y, and\ the\ values\ \alpha_i\ the\ change\ in\ Y'\ for\ each\ 1\ increment\ change\ in\ X_i$$

$$can\ be\ determined. For\ the\ calculation, refer\ to\ (Higgins, 2005).$$

Moreover, logistic regression generates the coefficient and the standard errors, with a formula to predict the logit function that gives the logarithm of the odds p/(1-p) transformation of the probability p of the presence of the response or the outcome as outlined below:

$$logit\ (p) = \ Y = Y_0 + \ \alpha_1 X_1 + \ \ \alpha_2 X_2 + \alpha_3 X_3 \ldots + \ \alpha_n X_n : \ (\mathbf{26}),$$

$where\ p\ is\ the\ probability\ of\ presence\ of\ a\ given\ outcome.$

$The\ logit\ transformation\ is\ defined\ as\ the\ logged\ of\ odds$:

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ characteristic}{probability\ of\ absence\ of\ characteristic}\ (\mathbf{27})$$

$$logit\ (p) = ln\ \frac{p}{1-p}\ (\mathbf{28})$$ (Schoonjans, 2016).

Multiple logistic regression has had different applications in various areas including diseases diagnosis and prognosis, etc., some of which are presented below.

- **Applications of the MLR Model**

The multiple logistic regression (MLR) technique is applied in this research as a probabilistic classifier to predict or determine the likelihood of distinguishing between patients with NPC1 disease and healthy controls or segregate between liver dysfunction disease subjects in experimental animals and controls or corresponding heterozygotes (Abbott and Carroll, 1984; Schoonjans, 2016; Taooka et al., 2014; Xia et al., 2013).

In this process, detection of the major biomarkers in these cases is important. Bearing in mind that in the present case, the original features are not the ones assessed by the MLR model, but scores vectors selected by the PCA. However, the resulting biomarkers in terms of original features can be detected given that the PCs are linear combinations of these original variables. The MLR pseudo-code is given below.

- **Pseudo-code: Adapted MLR**

Frame 5

1. *Select Dataset = {Plasma , Liver}*
2. *Select dataset Standardisation option = {*
3. *ON , OFF}*
4. *Select k folds = {5 , 10}*
5. *Select Sigmoid Function to represent the model*
6. *Split the dataset into training and testing set*
7. *Add an incept W_0 to the training dataset*
8. *Define threshold value*
9. *Get the Log likelihood value*
10. *Define the weight $wi = wi - \alpha\, df/dw^i\ [f(x,w)]$*
11. *Where I is the index of the i^{th} feature*
12. *Stop if difference between old and new log likelihood is below threshold f_0*
13. *Set initial weight w_o^i*
14. *$df/dw > f_0$*
15. *Where f(x, w) a cost function, and f_0 a threshold*
16. *While $df/dw > f_0$ is true, find df/dw, such that w^i. $df/dw < 0$*
17. *Then update weight using update rule*
18. *$w_i = w_i + \alpha\sum_k [y_k - p(y_k = 1|x_k, w)].x_k^i$*
19. *If w^i . $df/dw > 0$, then use updated weight w^i*
20. ***End** if*
21. ***End** while*
22. ***End** MLR*
23. *(**End** PCR)* (Smith, 2015)

Algorithm 5. MLR algorithm implementing principal component regression (PCR) for potential biomarker discovery for the intelligent tri-modelling techniques (ITMTs), with the end of the MLR corresponding also to the end of the PCR.

It should also be noted that handling high-dimensional features with only a small number of observations available is a complex issue in regression analysis. Regressing Y on the principal

components (PCs) works well, giving acceptable results (Artemiou and Li, 2009; Li, and Chiaromonte, 2007). Consequently, PCR as a model is visited next in this thesis.

4.4.6. Principal Components Regression (PCR) for Data Analysis

- **Theory**

In statistical analysis, principal component regression (PCR) is related to the regression analysis technique applied in the principal components analysis (PCA). This combination of principal components analysis and regression analysis is useful when there is not a clear understanding of the inter-relations between variables. In this case, hypotheses will be constructed on a small sample set, that will need to be verified later using a larger dataset (Massy, 1965).

Employing PCR in data analysis has been driven by the necessity to address collinearity issues in the dataset, especially amongst variables or predictors (Cook, 2007). Indeed, when too many predictors exist in a dataset, the task of designating the features responsible for the transformation at metabolomics level with regards to disease diagnosis or therapy study becomes awkward. Therefore, it is necessary to use principal component (PC) rather than scores vectors in order to facilitate our understanding of the system under study.

- **PCR Model Applications**

Different applications of principal component regression (PCR) have been reported in the literature. It has been utilised to understand the traits variance related to underlying genetic transformations. Some of the assessments applied were motivated on the necessity to determine whether multiple correlations could be found at the genetic level, which may affect and be detrimental for the genes as main biomarkers for trait variance (Schaid et al, 2002; Shen and Zhu, 2009). In another case, the assessment was based in developing a new model that was compared to existing model, with a view of improving the model's ability to predict progression (Abraham et al., 2016).

In this thesis, however, PCR is employed in two different stages that include the use of PCA to determine the main axes or the principal components (PCs) upon which the dependent variable Y is regressed. This step was followed by the multiple logistics regression (MLR)

model that applied its probabilistic prediction ability to detect the main biomarkers for NPC1 disease diagnosis.

Indeed, PCA strategy was employed in order to define the different PCs to be employed during the second phase that is MLR. In view of the concern raised about the PCs selection criteria not being fully rational, the PCs selected could be unrelated to the outcome (Wang and Abbott, 2008; Gauderman et al., 2007). Therefore, in the current thesis, the PC selection process was made more open, with the PCs chosen explaining up to 90% of the total variability, which ensures that only a low level of total variability was left unexplained. However, MLR employed different options that allowed the comparison of different combinations of variable selection, sample selection that resulted in the model's improvement in terms of biomarker discovery.

Nevertheless, the different features selected by the PCR technique have to be ranked in such a manner that is easy to detect their importance in understanding a disease aetiology. Three main ranking techniques were developed in this research and they are outlined below.

4.4.7. PCR and Features Ranking Techniques

- **Analysis Techniques**

In view of a strong correlation between the different features in the datasets, PCA was employed to select the PCs explaining the maximum variability in the original NPC1 disease datasets. In addition, the factor loadings and scores are used as new forms of data that will help in further analysis. The multiple logistic regression (MLR), the last algorithm of the intelligent tri-modelling techniques (ITMTs), was used to regress the PCs on the dichotomous variable Y, that takes the values of healthy (0) and diseased (+1). In this manner, the NPC1 disease features can be correlated to disease diagnosis. Therefore, a combination of the PCA and the MLR allows the implementation of a multivariate logistic regression model of the principal component scores vectors also known as principal component regression (PCR).
The factor scores are used as a new dataset to run the MLR and obtain a classification of the PCs with the regression coefficient r_i. Moreover, from the factor loadings, the coefficients of correlation noted α (in this thesis) between factors and features are obtained.

Two sets of coefficients are used to develop the biomarkers discovery techniques, as proposed in this thesis. This encompasses the heuristic method related to the determination of a rapid solution, which is appropriate for the time being, although it might not be the best. The sum of the product of the coefficients (SPC) refers to the calculation of the sum of the product of the coefficients of correlation and regression to deduce the contribution level of each PC. Finally, the exponential of the sum of the product of the coefficients refers to the use of the exponential value of the sum of the product of the coefficients calculated above.

- **Consideration:**

Different assumptions were made in the modelling section. They are all related to the rules applied for the selection of the key features. However, no clear-cut rule is set regarding which value (r, α, rα, etc.) to use and how to use them on completion of the PCR. After generating the PC scores vectors via PCA and the application of the MLR, there is no rational method to apply in the case of the multivariate analysis such key feature selections, and hence in defining the biomarkers. However, two coefficients are generated, which are the coefficient of regression and the correlation coefficient. Based on these coefficients, several combinations and rankings can be generated, and hence different features can be considered as being the most important, and therefore regarded as major biomarkers in NPC1 disease study.

The technique of rotating the PCs, which consists of selecting one PC as the main one and then monitoring the influence on the other features, that is monitoring the effect of this change. This technique is not readily applicable to this research since the main aim is to monitor and understand the correlation of the different features and their effect on one another. However, three alternative methods are presented, including the heuristic method, the sum of the product of the coefficients method, and finally the exponential sum of the product of the coefficient method. These methods are discussed in detail below.

- **Heuristic Method**

The heuristic method used is based on the fact that the algorithm searches for a suitable selection of the PCs that is sufficient, although the PCs selected are not necessarily the best ones. In this regard, priority is given to the coefficient of regression r, while the largest values of the coefficient of correlation α are chosen subsequently to perform the features ranking. For example, if PC_1, PC_2, and PC_3 are sufficient to explain the required level of variability, where

r_1, r_2, and r_3 are their coefficients of regression respectively and of value ordered $r_1 \geq r_2 \geq r_3$. This feature ranking combined three values that are the coefficient of regression (r_i), the coefficient of correlation (α_i) and the corresponding products $r_i\alpha_i$. This final value $r_i\alpha_i$ is the major value in the heuristic ranking, although r_i and α_i are carefully checked before validating the ranking based on the value of $r_i\alpha_i$. The final ranking will be x_3, x_2, and finally x_1 if the products $r_i\alpha_i$ are ranked as follow: $r_3\alpha_3 \geq r_2\alpha_2 \geq r_1\alpha_1$.

The heuristic approach is based on the researcher's ability to classify the features according to the values r, α and then rα, each in descending order. This approach allows us to make a choice, and attain a solution, although some choices (solutions) are not the best. The rationale behind this technique is in line with the definition of the heuristic method, which is finding a solution that will be primarily effective, although this may not represent / be the best possible solution (Martí and Reinelt, 2011).

- **Sum of the Product of the Coefficients (SPC) Method**

The sum of the product of the double coefficient method is applied in this research, and the rationale behind it can be established by the following mathematical reasoning based on equations **27** and **28** above:

In Multiple Logistic Regression, the equation below expresses that the patient is

*eithr from the NPC*1 *disease class or either from the contol class. This thesis used*

the code $+1$ *for being in the sick class and* 0 *for being in the healthy class.*

Let Y be a binary variable that can take values $+1$ *and* 0. *Y is defined as follow*:

$$Y = \begin{cases} +1 \; if \; individual \; has \; got \; the \; disease \\ 0 \;\; in \; case \; the \; individual \; is \; healthy \end{cases} \quad (\mathbf{29})$$

(Hosmer and Lemeshow, 2015; Salazar et al., 2012).

If we consider p as the probability of a patient having the NPC disease $(+1)$.

The odds is defined by the following ratio: *the probability of*

the presence over the probability of the absence by the equation:

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ disease}{probability\ of\ absence\ of\ disease} : (\mathbf{30})$$

The logit transformation is defined as the logged of the odds and is expressed

by the following equation: $\log(odds) = \ln\left(\frac{p}{1-p}\right)$ (Schoonjans, 2016): (**31**)

The regression equation can be expressed as follow: $Y = logit\ (p)$ *, with*

$logit\ (p) = r_1 F_1 + r_2 F_2 + r_3 F_3 + \cdots + r_k F_k + Y_o : (\mathbf{32})$, *where* $r_1, r_2, \ldots., r_n$ *are*

the regression coefficients (Abbott and Carroll, 1984). Equations (31)*and* (32)

give logit $(p) = \ln\left(\frac{p}{1-p}\right) = r_1 F_1 + r_2 F_2 + r_3 F_3 + \cdots + r_k F_k + Y_o : (\mathbf{33})$,

which is the equation of the logit transformation (Schoonjans, 2016).

By applying the function exponential to equation 33. *Let recall that*:

$\forall x \in [0, \infty[, \quad e^{\boldsymbol{ln}(x)} = \boldsymbol{x}$; *Thus*, $\quad e^{\ln\left(\frac{p}{1-p}\right)} = \frac{p}{1-p}$, *with* $\frac{p}{1-p} \geq 0$

$$e^{\ln\left(\frac{p}{1-p}\right)} = e^{(r_1 F_1 + r_2 F_2 + r_3 F_3 + \cdots + r_k F_k + Y_o)} : (33')$$

$\frac{p}{1-p} = e^{(r_1 F_1 + r_2 F_2 + r_3 F_3 + \cdots + r_k F_k + Y_o)} : (\mathbf{34})$, *by developing this equation* 34

the value of p can be determined, and expressed in MLR as:

$$p = \frac{e^{(r_1 F_1 + r_2 F_2 + r_3 F_3 + \cdots + r_k F_k + Y_o)}}{1 + e^{(r_1 F_1 + r_2 F_2 + r_3 F_3 + \cdots + r_k F_k + Y_o)}} : (\mathbf{35})$$

Using the function $\ln(X)$ *in equation* 34 *above, it comes that*:

$$\ln\left(\frac{p}{1-p}\right) = \ln\left(e^{(r_1 F_1 + r_2 F_2 + r_3 F_3 + \cdots + r_k F_k + Y_o)}\right) : (\mathbf{35}')$$

Recall: $\forall x \in \mathbb{R}, \ln e^x = x$; *Combining this with the logit transformation, it comes that*: $\ln\left(e^{(r_1 F_1 + r_2 F_2 + r_3 F_3 + \cdots + r_k F_k + Y_o)}\right) = r_1 F_1 + r_2 F_2 + r_3 F_3 + \cdots + r_k F_k + Y_o : (35'')$

from $35'$ *and* $35''$: $logit\ (p) = \ln\left(\frac{p}{1-p}\right) = r_1 F_1 + r_2 F_2 + \cdots + r_k F_k + Y_o = Y : (\mathbf{36})$;

The following term: $r_1 F_1 + r_2 F_2 + \cdots + r_k F_k$ *can be expressed as a sigma = sum*

Hence, $r_1 F_1 + r_2 F_2 + \cdots + r_k F_k = \sum_{j=1}^{j \le k} r_j F_j : (\mathbf{36}')$

Therefore, (36) *and* $(36')$ *give*: $Y = logit\ (p) = \ln\left(\frac{p}{1-p}\right) = \sum_{j=1}^{j \le k} r_j F_j + Y_o : (\mathbf{37})$

The equation of regression based on $logit(p)$ (Hosmer and Lemeshow, 2015; Salazar et al., 2012). *The coefficient* r_j *is the coefficient of regression of each PC to the determination of the probability of whether or not an individual is healthy or NPC*1 *disease carrier. In this expression,* F_j *are the principal components used and* Y_o *the intercept of the regression line.*

On the other hand, k *is the total number of principal components that have been selected by running the PCA and explaining the maximum variability in the datset.*

From the PCA factor loading, it is established that F_j is a linear combination of the original variables X_i. Thus it can be expressed as follow:

$F_i = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \cdots + \alpha_n X_n$: **(38)** – *Giving the following expression below* $F_i = \sum_{i=1}^{i \leq n} \alpha_i X_i$: **(39)**

The coefficient of correlation α_i is based on the relationship between the features and the PCs; where $i \leq n$ as it depends on the number of features used to define the F_i. By replacing (34) *by its value in* (32), *it comes that*:

$$logit\,(p) = \ln\left(\frac{p}{1-p}\right) = Y = \sum_{j=1}^{j \leq n} r_j \sum_{i=1}^{i \leq n} \alpha_{ji} X_i + Y_o : (\mathbf{40}) - \textit{that is}$$

$$Y = \sum_{j=1}^{j \leq n} r_j \left(\alpha_{j1} X_1 + \alpha_{j2} X_2 + \cdots + \alpha_{jn} X_n\right) + Y_o : (\mathbf{41}) - \textit{this gives the following}$$

$$Y = r_1\left[\left((\alpha_{11} X_1 + \alpha_{12} X_2 + \cdots + \alpha_{1n} X_n)\right)\right] + r_2\left[\left((\alpha_{21} X_1 + \alpha_{22} X_2 + \ldots + \alpha_{2n} X_n)\right)\right]$$

$$+ \cdots + r_m\left[\left((\alpha_{m1} X_1 + \alpha_{m2} X_2 + \cdots + \alpha_{mn} X_n)\right)\right] + Y_o : \textit{Using } X_i \textit{ as common factor}$$

$$\textit{it comes}: Y = (r_1\alpha_{11} + r_2\alpha_{21} + \cdots + r_m\alpha_{m1})\, X_1 + (r_1\alpha_{12} + r_2\alpha_{22} + \cdots + r_m\alpha_{m2})\, X_2$$

$$+ \cdots + (r_1\alpha_{1n} + r_2\alpha_{2n} + \cdots + r_m\alpha_{mn})\, X_n + Y_o = \ln\left(\frac{p}{1-p}\right)$$

It appears that the term $K_{ij} = \sum_{j=1}^{j \leq m} \sum_{i=1}^{i \leq n} r_j\, \alpha_{ji}$: **(42)** *represents the contributing coefficient of every single feature to the regression line Y. This sum will*

inform us of which one of these features contributed most greatly to the regres -sion line, i.e. which one is the most important marker in this regession analysis, or in this multivariate data analysis. Finally, Y can be expressed as follow:

$$Y = \ln\left(\frac{p}{1-p}\right) = \sum_{j=1}^{j\leq m}\sum_{i=1}^{i\leq n} r_j\, \alpha_{ji}\, X_i + Y_o \quad : (\mathbf{43}) - Where\ r_j\ is\ the\ coefficient\ of$$

regression and α_{ji} is the coefficient of correlation between features and PCs

m maximum number of PCs considered to explain the required variability level, n the maximum number of features in the original dataset and Y_o the intercept. The value p represents the probability of the individual having the disease.

The term K_{ij} that is a sum of product corresponds to the slot of the regression line. The value ofK_{ij} rerpresents the multiplicative factor by which each feature contributes to the regression line Y. Therefore, knowing the value of K_{ij} *will assist in feature rankings, and hence will help us to decide which features are the most important biomarkers in this research.*

Consideration:

Referring again to equation $(\mathbf{35})$, *the probability p can be rewritten as follow*

$$p = \frac{\exp\left(\sum_{j=1}^{j\leq m}\sum_{i=1}^{i\leq n} r_j\, \alpha_{ji}\, X_i + Y_o\right)}{1 + \exp\left(\sum_{j=1}^{j\leq m}\sum_{i=1}^{i\leq n} r_j\, \alpha_{ji}\, X_i + Y_o\right)}; given\ that\ \exp(a+b) = \exp(a)\exp(b)$$

$$p = \frac{\exp\left(\sum_{j=1}^{j\leq m}\sum_{i=1}^{i\leq n} r_j\,\alpha_{ji}\;X_i\right)\;\exp(\,Y_o)}{1+\mathrm{e}\left(\sum_{j=1}^{j\leq m}\sum_{i=1}^{i\leq n} r_j\,\alpha_{ji}\,X_i\right)\;\exp(\,Y_o)]}\;:(\mathbf{44});$$

hence by dividing by $\exp(\,Y_o)$ *it comes that* p *can be expressed as follow*

$$p = \frac{\exp\left(\sum_{j=1}^{j\leq m}\sum_{i=1}^{i\leq n} r_j\,\alpha_{ji}\,X_i\right)}{\exp(-Y_o)+\exp\left(\sum_{j=1}^{j\leq m}\sum_{i=1}^{i\leq n} r_j\,\alpha_{ji}\,X_i\right)}:(\mathbf{45})\;\textit{The term}\;\exp(-Y_o)\;\textit{is a constant.}$$

$$\textit{Therefore the term}\;\mathbf{exp}\left(\sum_{\boldsymbol{j=1}}^{\boldsymbol{j\leq m}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{i\leq n}}\boldsymbol{r_j\,\alpha_{ji}\,X_i}\right):(\mathbf{46})\;\textit{is the only term responsible for}$$

the variation in the contribution to the determination of the probability value p.

Practically, in the sum of the product of the coefficients (SPC) method, the values α_{ji} are obtained from the factor loadings, while MLR is run using the factor scores vectors as a new dataset to generate the coefficients of regression r_j. The coefficients of regression are then used in the factor loadings matrix and determined for every row, and the contribution of each explanatory variable by calculating the value of $\sum_{j=1}^{j\leq m}\sum_{i=1}^{i\leq n} r_j\,\alpha_{ji}$. The results of the technique are explained in more detail in the results section below.

- **Exponential of the Sum of the Product of the Coefficients (ESPC) Method.**

This alternative method uses the exponential value of the sum of the products of the two coefficients. The expression (10) established above justifies the rationale behind the (ESPC) method. The scalar generated allows us to classify features based on their exponential values. In this manner, negative coefficient will add some positive effect on the overall performance, because the function exponential being a positive and ascendant function therefore:

$$\forall\,(a,b)\in\mathbb{R}^2, if\;a\geq b\;\;then\;\exp(a)\geq\exp(b).$$

The values obtained and the techniques implementation, together with the result of the ranking are provided in the results section.

- **Consideration:**

To the best of the researcher's knowledge, no one has established such techniques with regard to PCR features ranking related to biomarkers discovery, or any such study based on the use of the principal components regression (PCR).

This extensive presentation of the different ranking techniques including the Heuristic, SPC, and the ESPC ones completed data multivariate analysis models employed. The results, especially the biomarkers discovered can now be interpreted and discussed with the overall techniques and models subjected to validation processes that confer more creditability to the whole intelligent tri-modelling technique and the findings.

4.5. Research Validation

4.5.1. Definitions

The term validation is a broad term that can have various interpretations depending on the area the research is applied to. Therefore, validity is said to be contextually related (Bapir, 2010). Validity is related to the extent to which an instrument can correctly perform the measurement it is supposed to make. In this respect, external validity is concerned with the generalizability of the measurement taken, while the internal validity assesses whether the measurement allows researcher to determine what he/she wants to know, what he/she has been researching on, etc. (Bapir, 2010). The validity of research results is of paramount importance, with researchers having to strive to make sure that their research method, including the findings, are valid and in line with appropriate standards (LeCompte and Goets, 1982; Research Rundowns, 2009).

The present research work's validity is tested and measured against standards using two main channels, including techniques employed during data analysis results, and the intelligent task technology fit models for research method alignment. A standard allows us to establish and define a unit of measurement, to give a reference for an instrument, a set of steps or 'norm' in a process in such a manner, that other instruments can be calibrated, or other process can tested against. In this thesis, the model developed the intelligent technology task fit model (ITTFM)

is an extension of the task technology fit model (TTFM) developed by Goodhue and Thompson in 1975, and is therefore validated against the TTFM used as a standard (Aguinaldo, 2004; Goodhue and Thompson, 1995). Based on this, the state of the ITTFM model attained a higher level of precision. Thus, a set of activities were defined to allow us to assess research processes reaching that of the standard in relation to further the use of CITs in the detection of biomarkers. In addition, a threshold was established to measure the standard level achieved by the model. For example, the performance rate was set 0.80. This included achieving the standard level in visualisation, classification, feature ranking, biomarkers discovery, etc. In this manner, the intelligent technology task fit model was validated, as outlined below.

4.5.2. Intelligent Technology Task Fit Model for Validation

The intelligent technology task fit model (ITTFM) was applied in this research study, but also as a research result and model validation technique. The alignment model defined is used to determine to what extent the data analysis technique at each stage is aligned with the task to be performed. In addition, if the performances are above the acceptance threshold, then combinations of criteria allow us to validate the research model developed and the results obtained. In contrast to the Goodhue and Thomson (1995) task technology model that was developed and based on users' acceptance of a model, their perception ease to use of a technology, and also the advantages and disadvantages of the technology in relation to the task to be performed, etc. (D'Ambra et al., 2013) to assess the model performance, the extension model applied here uses the CITs performances in terms of modelling to assess it. Therefore, the importance of visualisation, classification, ranking and biomarker discovery were used to judge the model fit for purpose against an established standard of acceptance. This indicates that the model is being assessed and validated against its practical performances, which are sketchy at best using valid performance metrics such as ROC AUC, with associated standard deviations and the 95% confidence intervals (Hummel et al., 2011; Kounev and Gorton, 2008).

The Table below assesses the ITTFM model using data analysis results, establishes the alignment level, and determines its fitness for this purpose based on the performance measured, and therefore that of the conditions to meet those of standards.

	Performance - Threshold Value	**Alignment Level**	**Intelligent-Fit**	**Meeting Standard – Conditions**
SVA	. Selection and ranking performance $r \geq 0.8 = r_0$	Very Good data visualisation and separation; hence alignment level $a = 0.85$	Good Performance Value Intelligent Model fit for purpose.	SVA allowing the visualisation of major potential biomarker(s) $r \geq 0.8$
SVM	. Iterative training and testing to tune the model with: Av. Gmean ≥ 0.82 Av. ROC AUC ≥ 0.83	Excellent rate of classification. Alignment level $a \geq 0.85$	Very Good Performance Intelligent Model fit for purpose.	SVM perfectly segregating between diseased and wild treated individuals. Gmean = 0.8 ROC AUC = 0.8
PCA	. Accurate selection of relevant PCs rate ≥ 0.8	Very Good PCs leading to potential biomarker discovery. Alignment level $a \geq 0.85$	Good Performance Intelligent Model fit for purpose.	PCA selecting relevant potential biomarkers Selection rate ≥ 0.8
PCR	. Modelling Quality: Av. Gmean ≥ 0.85 Av. ROC AUC ≥ 0.87 . Regression Quality: Av. Gmean ≥ 0.85 Av. ROC AUC ≥ 0.88	Very Good classifications rate. Alignment level $a \geq 0.85$ Very Good modelling rate. Alignment level $a \geq 0.85$	Very Good Performance Intelligent Model fit for purpose.	PCR detecting major potential biomarkers: ROC AUC ≥ 0.8

Table 3. ITTFM assessment based on the relative performance of each model, in relation to defined characteristics such as alignment level, performance threshold, and the intelligent model fitness for purpose. Av = average value, with the notion of standard being defined in

relation to the performance level established at 80%. The diagram below shows the relationship between the algorithms involved in this modelling.

4.5.3. Relationship between the Three Models Apply in the Research

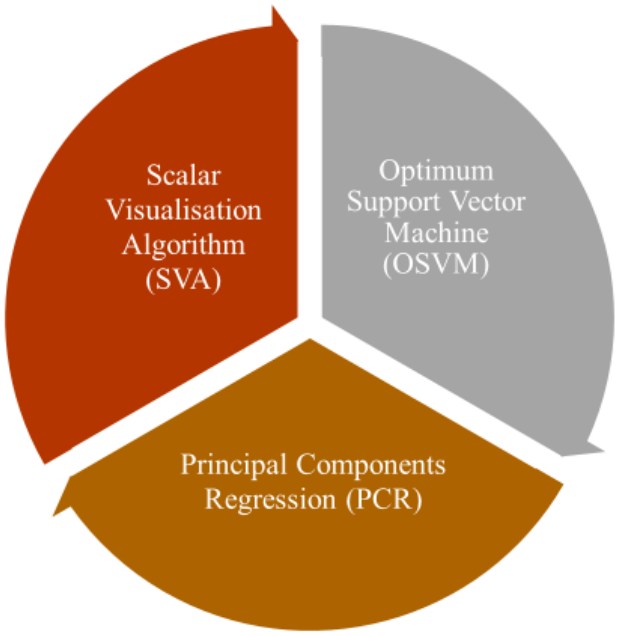

Figure 4. Relationship between the algorithms involved in the Intelligent Tri-Modelling Techniques (ITMTs)

- **Consideration**

Using the intelligent tri-modelling techniques (ITMTs), SVA allows to visualise the internal data structure highlighting area of defect in the samples analysed, while OSVM allows to establish a segregation between diseased and healthy subjects. Finally, PCR correlates the disease features to the disease development. All these will be further developed and detailed later in this thesis.

Nevertheless, the results of the analysis performed must be validated in order to provide more credit to the techniques and models developed as also presented in Figure 3 above. The next sub-section presents the use of the data analysis technique for validation purposes.

4.5.4. Data Analysis Techniques and Validation

Throughout this thesis, the two dataset analysis techniques results were validated against established results measurement standards using different methods and techniques, including:
1) The nearest neighbour features selection approach allowing the selection of the key features that can improve the OSVM overall performances, and they can be also potential biomarkers (Lazzarini and Bacardit, 2017; Wang et al., 2005).

2) The oversampling technique based on the ADASYN sampling approach, where synthetic data are generated in the minority class, which is harder to learn in order to balance the dataset and facilitate the learning stage (He et al., 2008). Indeed, learning algorithms can easily accommodate and learn large sets than they do with small sets.

3) The k folds stratified cross-validation approach allowing division of the main dataset into training and testing sets and repeat this technique k times across the dataset in order to use the whole dataset alternatively as training and testing sets (Scikit Learn, 2016; Schneider, 1997).

4) The intelligent technology task fit model (ITTFM) uses, a set of thresholds to classify the data analysis technique, especially the algorithm supporting it into one of the highest classes of alignment level, in order to perfectly align and ascertain that the model is fit for purpose. Indeed, the intelligent modelling process (IMP) could be used in combination with the ITTFM for example in thresholds setting.

5) The results of the analysis were based on validated performance metrics, including the ROC AUC and Gmean parameters. In addition, the standard deviation and the 95% confidence interval were provided for classification in the OSVM model. The detection of potential biomarkers that encompasses well-known biomarkers such as isoleucine, leucine, valine, Citrate, N-acetyl-4-O-acetylneuraminate, and glutamine etc., in the SVA model. Finally, potential biomarkers were detected using the PCR model, and that includes pyruvate, glutamate, taurine, hypotaurine, hippurate, adipate/n-butyrate, etc., for the NPC1 disease diagnosis. Other less well-known potential biomarkers were also detected, including adipate, hippurate, trigonelline, xanthine, etc. in the NPC1 liver dysfunction disease diagnosis.
The combination of these different techniques applied throughout this research has ensured that it and its findings are valid and meet standards.

4.5.5. Biomarkers Clinical Validation Technique

While the notion of validation in qualitative research is believed not to be a deterministic process, meaning that researchers do not have to demonstrate their methodology, their findings, etc., is valid or not, but should be perceived as a continuous process. However, quantitative research forces the researcher to go through the process of validation in a deterministic manner, i.e. researchers have to prove the validity of their work (Aguinaldo, 2004; Muijs, 2004). In this respect, it is understood that more accurate diagnosis and assessment of disease upstream should give more insight into disease development, and hence give clinical trials more weight in disease research, including useful diagnosis capability and important benefits regarding disease treatments. To be approved as a clinically valid test, potential biomarkers detected should undergo scrutiny with hundreds of samples being validated, and with a possibility to repeat and reproduce these tests, with high specificity, and sensitivity (Zhang et al., 2015).

Biomarkers discovered can be used as a clinical test in order to perform an early test for NPC1 disease diagnosis, while monitoring its progression. Relevant clinical tests to put in place in order to validate biomarkers discovered in this thesis, including adipate, methionine, hypotaurine, lysine, ornithine, phenylalanine, nicotinate, and xanthine, may now be based on miglustat therapy, the only available and valid treatment to the NPC1 disease approved by the American Food and Drugs Administration (Heron et al, 2012). Miglustat has shown to have a positive effect, i.e. a stabilising effect on many clinical biomarkers discovered in the study of the NPC disease (Patterson et al., 2007). Miglustat is an iminosugar, an inhibitor that plays a crucial role in the mice type of NPC disease, by reducing the accumulation of complex lipids such as gangliosides in the brain, and in such a manner that it delays the onset of the neurological dysfunction, and finally prolongs life expectancies for the diseased rodents. In humans, the administration of miglustat to NPC disease patients improves their condition considerably, including the amelioration of lipid trafficking, reduction of glycosphingolipid, reduction of pathological lipid storage, with an improvement in the endosomal uptake (Patterson et al., 2007).

In this respect, Patterson et al., (2007) used miglustat to treat NPC disease patients, postulating that the administration of the drug would slower the rate of decline of the patients in certain or all the biomarkers detected. Two main groups were used for this study, and these included those which were miglustat-treated and the care control groups. A group of 20 young NPC patients aged 12 years or more received a daily miglustat treatment of 3 x 200 mg, while another cohort of 9 children, same age group received a normal care. Both groups were monitored for one year. A third group of 12 children, younger than the precedent groups received also miglustat treatment, with an appropriate dose related to their body surface. In an extended study, the participants were all treated with miglustat for an additional year. The correlation of the disease progression and the velocity of the horizontal saccadic eye movement (HSEM) was studied. The first series of treatment revealed that the cohort receiving miglustat treatment had an improved HSEM compared to the ones who did not receive the miglustat treatment. In addition, the overall result showed an improvement in the quality of life of patients under miglustat therapy compared to the control group. Areas of improvement included swallowing, auditory acuity, and a slow decrease in the ambulatory stability for the patients aged 12 years or older treated with miglustat, where the drug tolerability was high (Patterson et al., 2007).

However, based on this study it is difficult to determine which biomarker is responsible for the amelioration of their life condition as mentioned above. Indeed, biomarkers GM2 and GM3 gangliosides concentration increased in diseased neurons, while the accumulation of cholesterol and glycosphingolipids is observed in many tissues (Patterson et al., 2007). Therefore, the present research can be improved by including the biomarkers as part of the variables studied and following the variation of their individual concentrations in both control and NPC disease patient groups in order to monitor response to the miglustat therapy on disease progression (Giese et al., 2015). Modification in the concentrations of biomarkers will signify their active involvement in miglustat therapy processes.

In another study, Probert et al., (2017) used ^{1}H NMR datasets containing information on miglustat-treated and untreated, and also heterozygotes and healthy control subjects. ^{1}H NMR data generated were analysed using advanced multivariate techniques. The regions of high-density-lipoprotein (HDL) triacylglycerols of the spectra was found to be effective for segregation between NPC1 disease and heterozygotes patients. In the same study conducted, the investigation was based this time based on NPC1 disease and healthy controls patients. It allowed to differentiate between NPC1 and healthy controls subjects, using ^{1}H NMR resonance

of lipoprotein triacylglycerol and isoleucine. The ^{1}H NMR profile of NPC1 patients might be related to the liver dysfunction associated with the disease. Potential biomarkers were identified due to the ^{1}H NMR signal of heterozygotes subjects compared to healthy control ones. Additionally, the increase level of amino acids observed might be ascribable to the liver parenchyma necrosis related to hepatic fibrosis (Probert et al., 2017).

Heron et al., (2012), used miglustat to treat three main categories of patients. These included those with early-infantile neurological onset, patients with late-neurological onset, and patients with juvenile neurological onset. The parameters included in their study are clinical parameters of neurological disease (NPC functional disability scale, epileptic seizure, cataplexy, video recording); whilst laboratory parameters encompasses liver function and haematology. Finally, other parameters included psychometric evaluation, hearing, and abdominal ultrasound. Descriptive statistics were used as the only way of analysis for the observation of the clinical changes. The results obtained showed that for early–infantile neurological disease patients, disease progression was slower; however, after 18 months the disability, the score increased as a sign of disease progression. Late–infantile neurological disease patients' disability scale score showed an overall stabilisation or improvement of the neurological parameter, with patients being freed from seizures after 5 months of miglustat therapy. For juvenile neurological onset patients, the disability score scale indicated mix outcomes, i.e. improvement then stabilisation for patients, and then use of an alternative therapy to control patient's cataplectic episodes. Overall, miglustat therapy produced more beneficial effect on late-juvenile patients than on early-juvenile ones (Heron et al., 2012). The introduction of surface metabolite levels following magnetic resonance spectroscopy (MRS), allowed researchers to compare the ratio of three main biomarkers, including N-acetyl-aspartate [NAA], creatine [Cr], and choline [Cho]. In addition, MRS enabled the identification of some abnormalities in the spectrum, with low NAA and/or high Cho producing an increased Cho/NAA ratio. Given that NAA is known to represent a biomarker of neuronal viability, its decreases in concentration are consistent with progressive neurodegenerative diseases progression, whilst the increase in Cho, which is believed to be a biomarker for membrane destruction or related to gliosis, is also consistent with the effect of miglustat on NPC disease patients (Heron et al., 2012). This study revealed also that the early the miglustat therapy started, the better the effect of the therapy on patients. For those who were administered the miglustat treatment late, such as 4.8 years after the disease onset, the condition worsened with this therapy (Heron et al., 2012; Pineda et al, 2009).

These last two studies show possible validation routes for the drug miglustat, together with the biomarkers included in the research performed here. Therefore, the ratio NAA/Cr, Cho/Cr, and Cho/NAA or only the variation in the concentration of each one of these biomarkers, could, at least in principle, be used to validate the biomarkers discovered. For example, the study did propose the high Cho level producing high Cho/NAA ratios as a marker for this therapeutic effect for late-infantile and forms of the disease manifestation, but not for the early-juvenile form. Similarly, the increase in concentration of hippurate [Hi] and adipate [Ad] in blood plasma samples of NPC1 disease patients when compare to wild type (WT), healthy individuals would indicate an active involvement of these biomarkers in miglustat therapy. This process can be used as a validation route for the biomarkers (hippurate and adipate) detected in this PhD thesis. The lack of available material did not allow the validation to be implemented in the present study, and consequently will be included in future research investigations.

4.5. Chapter Summary

The research design chapter includes an introductory section that highlights the main sub-sections of the chapter. Next is discussed the mixed experimental research method, which mentions the combination of fieldwork and the laboratory-based research involved in data collection. The biomedical data related to the NPC1 disease dataset collection was conducted by collaborative research institutes in Maryland, USA and Oxford, UK. The metabolomics data collection related to the NPC liver disease is then outlined. Further, data analysis tools were visited, giving detailed involvements of each of them in the development of the intelligent tri-modelling techniques. Finally, a short summary of the research design process concludes this chapter.

"The only way not to succeed is not to try"

Edward Teller

5. DATA ANALYSIS AND RESULTS

5.1. Introduction

The preceding Chapter (Research Design) explained the "mixed experimental" research method, which combined a fieldwork and a laboratory-based research methods. Computational Intelligent Technologies (CITs), including the scalar visualisation algorithm (SVA), the optimum support vector machine (OSVM) and principal component regression (PCR) involved in the data analysis formed a network of three major algorithms. They are applied in this multivariate analysis of metabolomics datasets based on the modelling of the Niemann-Pick Class 1 (NPC1) disease and NPC liver dysfunction disease (NPC LDD) datasets. Therefore, the modelling and data analysis techniques used are termed intelligent tri-modelling techniques (ITMTs). Beyond the modelling based on the numerical values (digits) contained in the datasets, the ITMTs enable an understanding of the underlying transformations occurring at molecular level in relation to the NPC1 disease and the NPC liver dysfunction disease (NPC LDD) diagnosis.

The first model developed, the SVA, provides some insights regarding a preliminary understanding of data structures. By creating a 'pointer' based on data values, the algorithm can assign those values to a set of colours held on a "colour look-up" table. In this manner, data dimensionality reduction which is, necessary for the mapping of data from higher to lower dimensionalities (2D) representation can be achieved, while coloured maps can highlight the internal data structure.

The second modelling phase is related to the use of the OSVM as a powerful algorithm for a deterministic classification that discriminates between diseased individuals and healthy controls in biomedical dataset analyse. Finally, to overcome the difficulty in the selection of

the exact feature (s) responsible for the underlying transformations at the molecular level, principal components regression (PCR) is applied as a model, allowing a probabilistic determination of potential biomarkers discovered in the different datasets explored.

5.2. Intelligent Tri-Modelling Techniques and the NPC1 Disease Dataset Multivariate Analysis

5.2.1. Introduction

The detection of biomarkers can be a possible solution to detect NPC1 disease's first symptomatic appearance, and perhaps also its progression. These tri-modelling techniques can be used to scrutinise the NPC1 disease dataset and determine biomarkers involved. Thus, the tri-modelling techniques applied includes a Matlab model for the NPC1 disease dataset visualisation, the optimum support vector machine (OSVM) for the dataset classification, and principal component regression (PCR) for potential biomarker detection.

Principal components regression (PCR) is implemented as a combination of the principal component analysis (PCA) and the multiple logistic regression (MLR) of the principal component scores vectors previously generated through PCA. For the OSVM and PCR, use of classification metrics such as the geometric means (Gmean), the receiver operating characteristic area under the curve (ROC AUC), confidence intervals (CIs), and the standard deviation(Std) will be used to assess the models' performance in order to validate the biomarkers discovered. The dataset involved this modelling process is below described.

5.2.2. Description of the NPC1 Disease Dataset

- **Description**

The NPC1 disease dataset encompasses 130 observations and 55 features (chemical shifts). They were obtained using NMR spectroscopy on blood samples collected from human subjects in hospital setting. The NMR spectrometry generates spectra amplitudes against radio frequency converted into chemical buckets expressed in ppm (part per million). Initially 133 observations were obtained and 3 of them were removed for different reasons. This includes

the rows with 0 values (1 observation) and rows with duplicated values (2 observations). The remaining 130 were carried forward for analysis. Amongst these, 71 were blood samples collected from healthy control subjects and the remaining 59 were NPC1 disease patients. The statistics obtained using XLSTAT are briefly shown below, the full set of this summary statistics is included in the appendix section.

- **Statistics**

Different statistics were generated from the XLSTAT software and are included in the following Table 4.

Summary statistics:							
Variable	Observations	with missing	without missing	Minimum	Maximum	Mean	td. deviatio
[0.68 .. 0.71]	130	0	130	0.000	0.003	0.001	0.000
[0.71 .. 0.73]	130	0	130	0.000	0.002	0.001	0.000
[0.73 .. 0.75]	130	0	130	0.000	0.002	0.001	0.000
[0.75 .. 0.77]	130	0	130	0.000	0.003	0.001	0.000
[0.77 .. 0.79]	130	0	130	0.000	0.004	0.002	0.001
[0.81-0.89]	130	0	130	0.095	0.187	0.132	0.018
[0.89 .. 0.95]	130	0	130	0.016	0.051	0.028	0.005
[0.95-1.06]	130	0	130	0.019	0.038	0.028	0.004
[1.13 .. 1.15]	130	0	130	0.000	0.006	0.002	0.001
[1.15 .. 1.17]	130	0	130	0.001	0.005	0.003	0.001

Table 4. Part of the summary statistics obtained from XLSTAT running PCA of the NPC1 disease dataset of human-model. The variables are chemical buckets with the means and standard deviation shown on the right hand-side. The very low standard deviation show that the data point are not widely spread out (for this partial summary statistics).

This visualisation of the internal data structure of the NPC1 disease (human-model) and the series of developmental processes involved in the data analysis are outlined below.

5.2.3. NPC1 Disease Data Visualisation

- **Scalar Visualisation Algorithm and the NPC1 Disease (Human-model) Dataset**

Matlab as a matrix-based computational language allows us to develop mathematical models for the analysis of datasets in general, and that of the current NPC1 dataset, in order to gain insights from the disease data (MathWorks, 2016). Two main options exist for Matlab to generate graphs. This includes the option of creating graphs through the Matlab toolbox, or entering in the command window the Matlab graphics command. The first option was chosen for its flexibility and allows Matlab to work as an unsupervised 'learner' that uses data presented to it, and model it in such a manner that the hidden underlying information related to the data features are made apparent. Therefore, Matlab, using the scalar visualisation algorithms (SVA), generates such plots. They allow the visualisation and targeting of important features involved in NPC1 disease (Human-model) diagnosis.

The colour mapping algorithm (CMA) maps scalar in the dataset to a specific colour in the colour look-up table, and hence defines a colour as a function of a scalar in the dataset (Komura, 2016; MathWorks, 2016a). Assorted colours are used to represent distinct regions in the dataset in such a manner that differences in the underlying data structure can be perceived and highlighted. However, the contouring or isosurfaces algorithm allows the construction of delimitation zones between different parts of the dataset. In this manner, the structural difference in the dataset can be identified. The application of these two algorithms generates the following graphs, in which the main features have been detected.

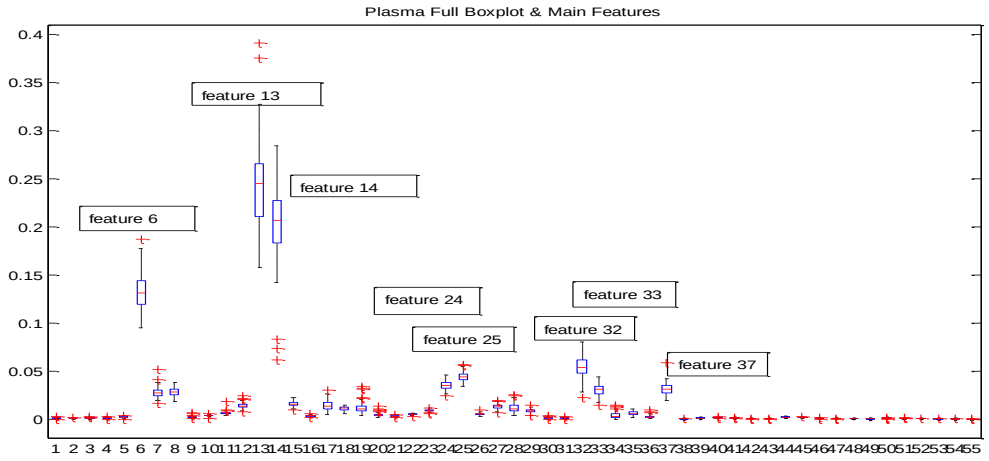

Figure 5. Plasma full boxplot showing features in order of importance according to the intensity level in the boxplot 13, 14, 6, 7, 8, 32, 25, 24, 33, 37, 28, 27, and 34 corresponding respectively to the following chemical shifts 1.21-1.31, 1.31-1.37, 0.81-0.89, 0.89-0.95, 0.95-1.03, 2.52-2.58, 2.03-2.09, 1.98-2.03, 2.68-2.74, 5.26-5.37, 2.34-2.39, 2.12-2.17, 2.85-2.88 ppm as the

most important biomarkers in the NPC1 disease (human-model) diagnosis. These correspond respectively to the following biomarkers hexacosanoate, isoleucine, leucine and valine as main ones.

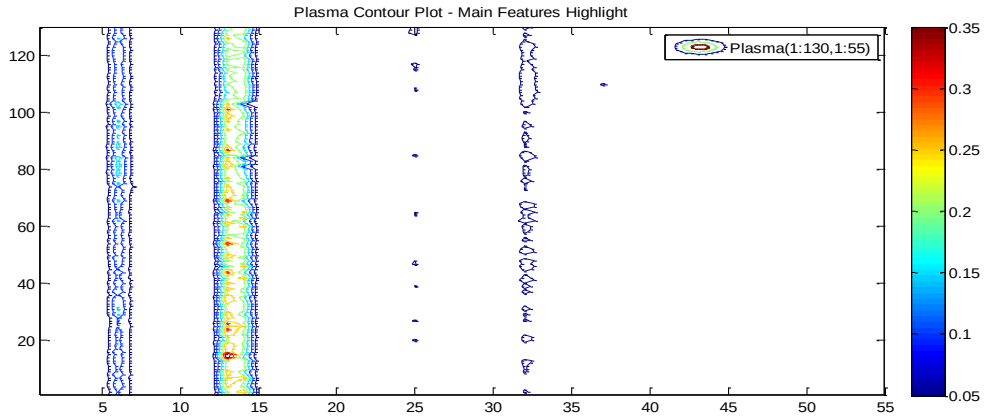

Figure 6. Plasma contour plot showing brighter colour for the main features 13, 14, 6, 7, 8, 32 and 25, 26. The values on the x axis correspond to the 55 features present in the plasma dataset. The corresponding biomolecules have ^{1}H NMR of chemical shift buckets 1.21-1.31, 1.31-1.37, 0.81-0.89, 0.89-0.95, 0.95-1.06, 2.52-2.58, 2.03-2.09, 2.09-2.12 ppm corresponding respectively to (R)-3-hydroxybutyrate, L-fucose, lactate; 3-hydroxyisovalerate, hexacosanoate, L-isoleucine, leucine, valine, Citrate, N-acetyl-4-O-acetylneuraminate, and glutamine respectively that are considered as potential biomarkers in NPC1 disease diagnosis.

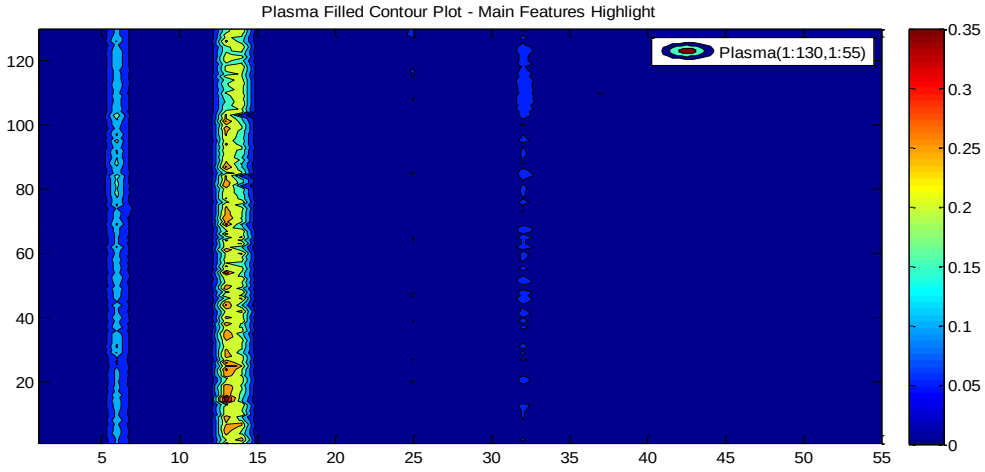

Figure 7. Plasma filled contour plot showing brighter colour for the features 13, 14, 6, 7, 8, 32, 25; since the values on the x axis correspond to the 55 features present in the plasma dataset.

The corresponding biomolecules have ^{1}H NMR of chemical shift buckets 1.21-1.31, 1.31-1.37, 0.81-0.89, 2.52-2.58, 2.03-2.09 ppm corresponding respectively to (R)-3-hydroxybutyrate, L-fucose, lactate; 3-hydroxyisovalerate, L-isoleucine, leucine, valine, Citrate, and N-acetyl-4-O-acetylneuraminic acid respectively that are considered as potential biomarkers in NPC1 disease (human model) diagnosis.

- **Consideration**

The SVA using the technique of look-up table transforms the peak amplitude values into colours from bleu (low) to orange (high).

- **Results and the Data Analysis**

Results of the scalar visualisation process have been grouped in the following table in order to highlight the main features from use of the different techniques applied to the NPC1 disease dataset.

Data Visualisation Technique	*Features Counting in Descending Order of Importance*	*Corresponding Features (Chemical Shift/ppm)*	*Corresponding Features /Biomolecules Name*
Boxplot	13, 14, 6, 32, 25, 24, 33, 37, 28, 27, 34	[1.21…1.31] [1.31...1.37] [0.81…0.89] [2.52…2.58] [2.03…2.09] [1.98…2.03] [2.68…2.74] [5.26…5.37] [2.34…2.39] [2.12…2.17] [2.85…2.88]	3-Hydroxybutyrate; L-Fucose; 3-Hydroxyisovalerate; Lactate; Hexacosanoate; N-acetylaspartate; N-acetylneuraminate; N-acetyl-X-CH_3; 2-hydroxyglutarate; Citrate; α-Glucose; Allantoate; Pyruvate; Glutamate; Methionine; Glutamine; Trimethylamine
Plot	13, 14, 6, 32, 25, 24, 33, 37	[1.21…31] [1.31…1.37] [0.81…0.89] [2.52…2.58] [2.03…2.09] [1.98…2.03] [2.68…2.74] [5.26…5.37]	3-Hydroxybutyrate; L-Fucose; 3-Hydroxyisovalerate; Lactate; Hexacosanoate; N-acetylaspartate; N-acetylneuraminate; N-acetyl-X-

			CH_3; 2-hydroxyglutarate; Citrate; α-Glucose; Allantoate;
Data Scale Plot	13, 14, 6, 32, 25, 24, 33, 37, 28, 27, 34	[1.21…1.31] [1.31…1.37] [0.81…0.89] [2.52…2.58] [2.03…2.09] [1.98…2.03] [2.68…2.74] [5.26…5.37] [2.34…2.39] [2.12…2.17] [2.85…2.88]	3-Hydroxybutyrate; L-Fucose; 3-Hydroxyisovalerate; Lactate; Hexacosanoate; N-acetylaspartate; N-acetylneuraminate; N-acetyl-X-CH_3; 2-hydroxyglutarate; Citrate; α-Glucose; Allantoate; Pyruvate; Glutamate; Methionine; Glutamine; Trimethylamine
Contour Filled Plot	13, 14, 6, 32, 25, 24, 28, 27, 34	[1.21…1.31] [1.31…1.37] [0.81…0.89] [2.52…2.58] [2.03…2.09] [1.98…2.03] [2.34…2.39] [2.12…2.17] [2.85…2.88]	3-Hydroxybutyrate; L-Fucose; 3-Hydroxyisovalerate; Lactate; Hexacosanoate; N-acetylaspartate; N-acetylneuraminate; N-acetyl-X-CH_3; 2-hydroxyglutarate; Pyruvate; Glutamate; Methionine; Glutamine; Trimethylamine
Contour Plot	13, 14, 6, 7, 32, 25, 33	[1.21…1.31] [1.31…1.37] [0.81…0.89] [0.89…0.95] [2.52…2.58] [2.03…2.09]	3-Hydroxybutyrate; L-Fucose; 3-Hydroxyisovalerate; Lactate; Hexacosanoate; isoleucine; N-acetylaspartate; N-acetylneuraminate; N-acetyl-X-CH_3

Table 5. Patterns of features detected in the NPC1 disease dataset, independently of the data visualisation plotting technique – Chemical shifts and potential biomarkers names.

- **Considerations**

Two important plotting techniques were applied in the NPC1 disease dataset. The colour map algorithm (CMA) is based on the definition of a function between the scalars values in the data matrix and used as index to locate the colours stored in the colour look-up table. This technique allowed the mapping and plotting of the 76 features resonances for 159 samples collected during the fieldwork and the laboratory-based research. The plots generated include plot, boxplot, data scale plots, and these enabled consistent detection of the same features, i.e. 13,

14, and 6 corresponding to the following biomolecules 3-hydroxybutyrate; L-fucose; 3-hydroxyisovalerate; lactate; hexacosanoate as the most important ones for NPC1 disease (human-model) diagnosis. Other features were found to be important during the data visualisation process, including features 32, 25, 24, 33, 37, 28, 27, and 34 corresponding to the biomolecules listed below N-acetylaspartate; N-acetylneuraminate; N-acetyl-X-CH_3; 2-hydroxyglutarate; Citrate; α-Glucose; allantoate; pyruvate; glutamate; methionine; glutamine. The colour coded technique applied is from blue (low) to red (high) peak intensities.

However, the contouring algorithm creates boundaries between different regions of the data characterised by constant amplitude or intensity values. The plots generated through this technique selected features 13, 14, and 6 as the most important ones for NPC1 disease. Other features, including 32, 25, 24, 33, 37, 28, 27, 34 corresponding to the biomolecules listed below N-acetylaspartate; N-acetylneuraminate; N-acetyl-X-CH_3; 2-hydroxyglutarate; Citrate; α-Glucose; allantoate; pyruvate; glutamate; methionine; glutamine were considered important, but to a lesser degree.

Furthermore, feature 7 detected by the contour plot as an important feature was classified by the boxplot as a zero none deterministic feature in the dataset. These results related to the importance of the features mentioned above has to be verified using alternative data analysis techniques such OSVM and PCR. Thus, OSVM will be employed in section 5.2.3.

5.2.4. Optimum Support Vector Machine (OSVM) and the NPC1 Disease Dataset Classification

The optimum support vector machine is the second model developed in this intelligent tri-modelling technique applied. Used as a deterministic binary classifier, it employs a hyperplane to segregate between two classes of individuals, typically NPC1 patients and the wild treated individuals (Fu et al., 2016). Below, the technique employed is dissected in order to reveal the particularity of the OSVM strategy.

- **Data analysis techniques**

In the present research work, the OSVM is used for a binary or dichotomous classification of the NPC1 disease dataset based on the disease's potential biomarker. In order to detect major discriminants in the NPC1 disease dataset that might inform us of the disease's early development and progression, a setting is put in place to enable the OSVM classifier to segregate between the two classes of individuals diseased and healthy. The setting of this OSVM classifier involves making an optimal choice in terms of feature selection. The combination of choice should render the OSVM classifier to be effective and efficient with respect to selection of the major biomarkers.

The first choice made includes the selection of the best kernel function to be used in the data analysis stage, i.e. choosing between linear are RBF kernel. Although the RBF kernel has been presented in several studies as out-performing the linear kernel and many other classification algorithms (Maji et al., 2013; Sharma et al., 2011); in this current study linear and RBF kernel functions are used in order to compare their performances from several runs on the actual NPC1 disease dataset. However, combinations and optimisations of the penalty cost for misclassification noted in this study costC, the kernel function parameter gamma (γ) through the grid search algorithm was found to profoundly influence the model's generalisability (Akay, 2009). Similar techniques employed in previous research have subsequently allowed reduction of the computational cost by using a heuristic search technique (Boardman and Trappenberg, 2006; Fu et al., 2016; Maji et al., 2013; Mountrakis et al., 2011).

The second choice is related to the 'tuning' of the OSVM classifier through costC and gamma selection. Gamma is considered as the inverse of the influence of sample(s) selected by the model as a support vector, which can be far from or close to the decision boundary respectively for low or high values of gamma. Therefore, it determines the generalisation and the complexity of the SVM built, while allowing the monitoring and the fitting of a separation plan between the different classes explored (Boser et al., 1992; David et al., 2016; Pedregosa et al., 2011; Verplancke et al., 2008).

Notwithstanding, costC is known as a 'penalty' cost for misclassifying samples, i.e. it is a trade-off between the optimisation of the margin width and minimising sample misclassification rates (Boardman and Trappenberg, 2006). Misclassifications affect the decision surface, with low

values of costC indicating a smoother decision surface, while higher values suggest more constraint on the classification algorithm (see 4.4.3 in CH4). This includes, for example, attempts to perfectly classify all the training samples, giving rise to small levels of bias (Cortes and Vapnik, 1995; Pedregosa et al., 2011; David et al., 2016).

For the models' implementation, values of costC and gamma were determined through a grid search technique based on a heuristic one. The heuristic allows us to gain a solution for a given problem although it is not always the best solution, or it can simply approximate the exact solution to the problem (Boardman and Trappenberg, 2006; Martí and Reinelt, 2011). This process handled by the grid search algorithm reduces the searching space in the purpose of finding local minima (Woodford and Phillips, 2011). Hence, based on the precedent information, the algorithm approximates the result and decides on the next step that leads to the identification of the local maximum or the optimal solution.

The third choice involves the selection of the best classifier performance metrics. The choice is made between the area under the curve of the receiver operating characteristics (AUC ROC), and the geometric mean (Gmean).

The fourth choice is related to the determination of validating instance. The performance metrics are compared against each other, together with the standard deviation and confidence intervals related to those values. Therefore, the standard deviation will inform on how far from the mean the data values are spread. Confidence intervals for each choice, which will inform us regarding whether or not the values obtained fall within the 95% confidence intervals.

The fifth set of choice, which is related to the type of features and sample selection techniques implemented, are defined and are related respectively to the repeated k folds stratified cross-validation (RSCV) techniques and the nearest neighbour features selection (NNFS). The RSCV is a cross validation technique which allows the introduction, in every single sub-set of the current dataset, a fair representation of each class involved in the OSVM classification. It generates a more reliable estimate than running a single cross-validation technique (Kim, 2009). However, the ADASYN oversampling technique is implemented to balance the NPC1 dataset. By generating synthetic data samples in the minority class, it aims to obtain approximately the same number of diseased patients and wild treated cohorts. This allows the OSVM classifier to make an improved prediction, as well avoiding any bias in favour of the

majority class (Blagus and Lusa, 2012; He et al., 2008). Overall, the 'tuning' of these different parameters is a very important part of the optimisation strategy adopted in order to improve the overall OSVM classifier performances when compared to similar SVMs using different settings (Kim, 2009).

- **Results and analysis**

Different combinations of the OSVM classifier setting were used in generating the following results. This includes, using the nearest neighbour samples for oversampling (NNO), with the ADASYN algorithm. The repeated stratified K folds cross-validation simply labelled K- folds used (KFU) and the number of repeated stratified cross-validation (RSCV). The proportion of features used (POFU) for the reliefF based features selection technique is stronger in feature interactions and does not really function on the basis of heuristics (Nguyen and de la Torre, 2010). Finally, standardisation of the dataset could be set "ON" (SON) or set "OFF" (SOFF). The results of these combinations related to our ^{1}H NMR plasma dataset are reported below.

- **OSVM Results Employing the Kernel RBF and Gmean Performance Metrics**

The general condition applicable is as follows: NNO = 10; KFU = 10; RSCV = 10; NNFS = 10; POFU = 1; SOFF --- RBF and Gmean --- The values in the 1st column (blue) relate to that general condition. Changes are included in the changed variables section.

Changed Variable(s)	NNO=10	NNO=15	NNFS=15	POFU=0.5	POFU=0.75	KFU=5
Gamma	**11.65952**	4.088607	3.267071	15.67023	15.50455	2.81658
CostC	**8655.016**	5410.832	8837.593	547.7823	1515.231	4064.99
Average Gmean	0.884739	0.885154	0.886074	**0.886129**	0.884972	0.88474
Standard Deviation	0.012223	0.013046	0.01093	**0.012036**	0.013471	0.01408
95% Confidence Interval	**±0.02396**	±0.02557	±0.02142	**±0.023591**	±0.026404	±0.0276

Table 6. Modification of variables and the corresponding increases in OSVM classifier performances with the percentage of folds used (POFU = 0.5) giving the best average value of Gmean = 0.886129. The average values of these average Gmean = 0.88530

- **Results**:

The OSVM classifier performances using RBF kernel and the Gmean as performance metrics shows that an increase in oversampling, the number of nearest neighbour features selected, the proportion of features used, the number of K folds used and the number of repeated stratified cross-validations improves OSVM classifier performance.

These results suggest that increasing selected parameters, including NNO, KFU, RSCV, NNFS, and POFU increases the overall classifier power in discriminating between the NPC1 disease patients and the wild treated individuals, with higher accuracies indicated by standard deviation and 95% confidence interval reductions.

- **Results and Analysis of OSVM Performances for Linear Kernel and Gmean Performance Metrics**

The general conditions applicable are: NNO = 10; KFU = 10; RSCV = 10; NNFS = 10; POFU = 1; SOFF --- Linear and Gmean --- The values in the 1st column (blue) relate to the general condition. Modifications are included in the modified variables section.

Modified Variable (s)	NNO=10	NNO=20	NNFS=15	POFU=0.75	KFU=5
Average Gamma	0	0	0	0	0
Average CostC	16411.91	18577.03	15715.29	16666.77	16411.91
Average Gmean	0.875935	0.87298	**0.877246**	0.864075	0.86569
Standard Deviation	0.014592	0.013232	0.014508	0.012844	0.01609
95% Confidence Interval	±0.028601	±0.025934	0.028435	±0.025174	±0.031536

Table 7. Modifications of variables and corresponding increases in OSVM classifier performances and the average Gmean = 0.8711852, with the increase in the NNFS giving the best performance value average Gmean = 0.877246.

- **Results**

The OSVM Classifier Performances obtained using Linear kernel and the Gmean as performance metrics showed that increasing the number of nearest neighbour features selected, the proportion of features used, the number of K folds used and the number of times the stratified cross-validation is repeated, ultimately improves OSVM classifier performances. This result suggests that increasing the different parameters including NNO, KFU, RSCV, NNFS, and also the POFU, increases the overall classifier power in discriminating between the NPC1 disease carriers and the wild treated individuals, with higher accuracies given that the standard deviation and the confidence intervals decrease.

For this model, the increase oversampling does not show increases in OSVM classifier performances. This may be ascribable to several factors, such as over-fitting that renders the classifier to model noise instead of actual data (Kotsiantis et al., 2006; Lin and Chen, 2013; Tang et al., 2009).

- **Results and Analysis of the OSVM Performance Based on Linear Kernel and the AUC ROC Performance Metrics**

The general conditions applicable (values in the 1st column Table 7 below – in blue) are: NNO = 10; KFU = 10; RSCV = 10; NNFS = 10; POFU = 1; SOFF. Modified variables are included in the modified variables section.

Modified Variable (s)	NNO=10	NNO=15	NNFS=15	POFU=0.5	KFU=5
Average Gamma	0	0	0	0	0
Average CostC	21673.04	22804.12	23142.06	5873.269	22541.95
Average AUC ROC	0.939407	**0.93997**	0.939514	0.933813	0.932763
Standard Deviation	0.005404	0.005456	0.006524	0.005194	0.008348
95% Confidence Interval	±0.010592	±0.010693	±0.012787	±0.010181	±0.016362

Table 8. Modifications of variables and corresponding increases of the OSVM classifier performances, with the average value of the AUC ROC value being 0.9371 with the increase in the NNO giving the best performance value average Gmean = 0.93997.

- **Results**

OSVM Classifier Performances using the linear kernel and the AUC ROC as performance metrics. This shows that increasing the oversampling, the number of nearest neighbour features selected, the proportion of features used, and the number of k folds used in the cross-validation strategy employed improves OSVM performance. This result suggest that increasing the differing parameters, including NNO, KFU, RSCV, NNFS, and POFU increases the classifier power in segregating between the NPC1 disease carriers and the wild treated individuals (human model), with higher accuracies deriving from the corresponding standard deviation and the confidence interval decreases.

- **Results and Analysis of the OSVM Performance Based on the RBF Kernel and the AUC ROC Performance Metrics**

The general conditions applicable (values in the 1st column Table 8 below – in blue) are:
NNO = 10; KFU = 10; RSCV = 10; NNFS = 10; POFU = 1; SOFF --- RBF and AUC ROC ----. Modified variables are included in the changed variables section.

Changed Variable (s)	NNO=10	NNO=15	NNFS=15	POFU=0.75	KFU=5
Average Gamma	21.90145	14.68029	22.83431	2.026036	5.563069
Average CostC	5598.358	578.5882	521.4192	11631.33	4963.132
Average AUC ROC	0.944326	0.944657	**0.94784**	0.940831	0.94188
Standard Deviation	0.005268	0.005126	0.006619	0.008974	0.007157
95% Confidence Interval	±0.010326	±0.010048	±0.012973	±0.017588	±0.014028

Table 9. Modifications of variables and corresponding increases in the OSVM classifier performances. The average value of the AUC ROC was 0.9439, where NNFS produces the best average Gmean = 0.94784.

Results:

The OSVM Classifier Performances obtained using the RBF kernel and the AUC ROC strategies as performance metrics shows that increasing the oversampling, the number of nearest neighbour features selected, the proportion of features used, the number of K folds

employed, and the number of times the stratified cross-validation was repeated, improved OSVM performances.

The results acquired suggested that increasing the different parameters including NNO, KFU, RSCV, NNFS, and POFU, increased, the classifier power in segregating between the NPC1 disease carriers and the wild treated individuals, with higher accuracies, as noted from decreases in the standard deviation and confidence intervals in this human model study.

- **Results and analysis of the OSVM performance: Comparison of the functions applied and performance metrics**

The general condition applicable is as follow: NNO = 10; KFU = 10; RSCV = 10; NNFS = 10; POFU = 1; SOFF with --- Linear and Gmean; Linear and AUC ROC; RBF and Gmean; RBF and AUC ROC --- Changes are included in the modified variables section.

Modified Variable(s)	Linear and Gmean	Linear and AUC ROC	RBF and Gmean	RBF and AUC ROC
Average Gamma	0	0	11.65952	21.90145
Average CostC	16411.91	21673.04	8655.016	5598.358
Average Gmean /AUC ROC	0.875935	0.939974	0.884739	0.944326
Standard Deviation	0.014592	0.005097	0.012223	0.005268
95% Confidence Interval	±0.028601	±0.00999	±0.023957	±0.010326

Table 10. Modifications of variables and corresponding increases of the OSVM classifier performance - The AUC ROC outperformed with respect to Gmean values while RBF application gave an improved performance over that of the linear kernel technique.

- **Results**

The OSVM classifier performances by arising via modifications of linear to RBF kernels, and the Gmean to AUC ROC performance metrics obtained showed an increase in the OSVM classifier performance. These results suggest that combination the RBF kernel with AUC ROC approaches increases the classifier power in segregating between the NPC1 disease patients

and the wild treated individuals, with higher accuracy, and a notable decrease in standard deviation and associated 95% confidence interval (human-model).

It should to be noted that the increase in the discriminating power is improved using RBF and AUC ROC respectively as a kernel function than using the linear function and Gmean approach, which is in line with previous research conducted (Martí and Reinelt, 2011; Tang et al., 2009).

- **Consideration:**

With regards to the tuning of the different parameters involved in the optimum support vector machine (OSVM) classification, all the parameters (NNO, NNFS, POFU, RSCV, and KFU) increase the classifier's performances. Therefore, the OSVM becomes more powerful in discriminating between the NPC1 diseased patients and wild treated individuals. In addition, it was established that the RBF kernel over-performed the linear kernel, while the AUC-ROC provides an improved result over that of the Gmean. This final result will be used in the remaining data classification, and also for reasons of conciseness. Thus, the only function used will be the RBF kernel, while the performance metric used will be the AUC-ROC one. Given that the results achieved by the OSVM classifier that featuring, principal component regression is next to follow.

5.2.5. Principal Component Regression for Biomarkers Discovery in NPC1 Disease Diagnosis

Principal component regression is being implemented in two stages, and these include principal component analysis in order to select PC scores vectors and regressing the response variable (Y) axis on these PCs. Thus, due to a very strong correlation between the features involved in the NPC1 disease analysis, it appeared necessary to used uncorrelated features that will clearly attribute features level of involvement on the aforementioned disease diagnosis.

- **Principal component analysis (PCA) for the selection of the principal components (PCs) for NPC1 disease diagnosis**

- **Analysis techniques**

PCA is used here as a data analysis tool since it allows a dimensionality reduction from higher to much lower ones, where only a few of those PCs will be sufficient to explain the variability observed in the dataset. Therefore, by applying PCA to the original NPC1 disease dataset, this research aims at defining new features, i.e. principal components (PCs) that are uncorrelated in order to determine their importance in NPC1 disease diagnosis. Once the PCs defined, defined, it is easier to return to the original dataset and determine the original features' roles in the NPC1 disease aetiology and diagnosis. The analysis undertaken was based on certain numbers of considerations, the main one based on the importance of the eigenvectors and their implications. The assumption made is that given that none of the factors are simply predominant explaining the variability in the original dataset. Therefore, many of the PCs were considered to explain the maximum variability in the NPC1 disease (human-model) dataset. In addition, different levels of variability were considered in this analysis. Therefore, 90% variability was first considered, i.e. F1, F2, F3,…, F13, and F14 were necessary to explain this level of variability in the NPC1 dataset.

Thereafter, the 80% of variability was considered which necessitated, in turn, the involvement of the principal components F1, F2, F3,…, F8 and F9. Next was the case of the explanation of 70% of the variability in the NPC1 disease dataset, which necessitated the PCs F1, F2, F3, F4, and F5. Finally, 60% variability was considered, and which only necessitated the PCs F1, F2, and F3 to explain this level of variability in the dataset.

- **PCA results related to the NPC1 disease diagnosis**

The results generated are presented below, and include factor scores vectors, eigenvalues, and finally factor loadings that provides the influence and or contribution of each feature in corresponding PCs. These values will be employed subsequently in the selection of the main markers required for NPC1 disease diagnosis.

- **Eigenvalues**

Principal Components	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
Eigenvalue	16.431	10.680	4.832	4.168	2.782	1.680	1.608	1.277	1.170	1.075	0.985	0.890	0.831	0.652
Variability (%)	29.875	19.419	8.785	7.578	5.059	3.055	2.923	2.321	2.127	1.955	1.792	1.618	1.511	1.186
Cumulative %	29.875	49.293	58.079	65.656	70.715	73.770	76.693	79.014	81.141	83.096	84.887	86.505	88.017	89.203

Table 11. Eigenvalues from principal components 1 (F1) to principal components 14 (F14), where F1<=>F14 explains more than 89% variability, meaning that these 14 PCs are enough to explain maximum variability in the dataset.

- **Factor Loadings**

Chemical Buckets	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
[0.68 .. 0.71]	-0.105	0.805	0.272	-0.343	0.174	0.105	-0.081	0.051	-0.073	0.007	0.022	-0.103	-0.086	0.139
[0.71 .. 0.73]	0.052	0.909	0.163	-0.244	0.147	0.064	-0.063	0.008	-0.078	0.034	0.061	-0.057	-0.013	0.094
[0.73 .. 0.75]	0.141	0.925	0.259	-0.152	0.059	0.027	-0.043	0.025	-0.043	0.017	0.025	-0.050	0.004	0.056
[0.75 .. 0.77]	0.159	0.922	0.277	-0.138	0.036	0.007	-0.025	0.016	-0.050	0.021	-0.003	-0.055	0.000	0.069
[0.77 .. 0.79]	0.114	0.941	0.154	-0.169	0.111	-0.026	-0.053	0.002	-0.044	0.023	-0.010	-0.046	-0.008	0.052
[0.81 .. 0.89]	0.037	0.161	0.581	0.601	-0.172	-0.257	0.000	0.114	-0.124	-0.042	-0.195	-0.042	0.057	-0.042
[0.89 .. 0.95]	-0.087	-0.056	-0.610	0.078	0.629	-0.174	0.000	-0.074	0.046	0.160	0.148	-0.006	-0.013	-0.047
[0.95 .. 1.06]	0.555	0.243	-0.312	0.260	0.030	0.085	-0.047	-0.375	0.048	0.306	0.150	-0.104	-0.089	-0.205
[1.13 .. 1.15]	0.138	0.563	0.300	-0.302	0.334	0.159	0.162	0.086	-0.104	0.022	0.272	-0.137	-0.021	0.069
[1.15 .. 1.17]	0.137	0.865	0.186	-0.162	0.246	0.061	0.016	-0.020	0.092	0.024	0.070	0.005	-0.006	0.062

Table 12. Partial visualisation of contributions of the NPC1 disease features to each one of the factors (PCs), where positive values are showing important contribution to the corresponding factor. The chemical buckets correspond to the features and are the biomolecules resonance frequency interval expressed in part per million (ppm).

- **Factor scores**

Observations	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
Obs1	-0.566	-2.816	-0.299	-2.194	1.548	1.377	0.373	0.482	0.969	3.066	0.130	2.107	-0.179	0.194
Obs2	1.792	-1.611	-4.272	-1.779	1.358	-1.701	0.793	-0.774	-1.115	0.139	-0.635	0.064	0.300	-0.019
Obs3	0.066	-2.933	0.119	-1.895	-0.294	0.511	0.338	-1.603	-0.846	-0.240	0.061	0.274	0.433	0.309
Obs4	1.057	-2.154	0.179	0.445	0.672	-0.716	0.410	-0.769	-0.021	0.429	0.247	-0.703	-0.343	-0.725
Obs5	-1.680	-2.818	1.183	-1.070	-0.157	1.470	1.023	-0.868	-0.127	-0.245	-0.102	-0.072	0.813	0.671
Obs6	-4.669	-0.289	0.149	-1.912	-0.869	-0.352	0.539	-2.523	-0.295	0.011	0.112	-0.032	0.819	0.894
Obs7	-3.849	2.158	-1.911	-1.626	-0.568	1.353	0.572	-0.095	0.330	-0.868	-0.539	-0.412	0.799	-0.254
Obs8	-0.968	-0.608	1.132	-0.445	0.566	-0.966	-0.216	-0.077	0.457	-0.596	-0.870	-0.598	-0.001	0.232
Obs9	-0.051	-3.845	-2.928	2.824	2.408	1.418	-0.980	0.624	1.849	1.208	-0.274	-0.010	1.374	0.211
Obs10	3.259	-0.402	0.441	-2.988	-1.142	0.190	-1.006	-2.655	1.320	0.317	0.306	0.879	1.473	1.574

Table 13. Partial visualisation of the factor scores related to the scores of the contribution of each observation to the factors/principal components of the NPC1 disease (human-model) features. The high positive values expressing important contribution to the corresponding factor. For instance, observation 10 is making a big contribution to the factor F1.

- **Biplots:**

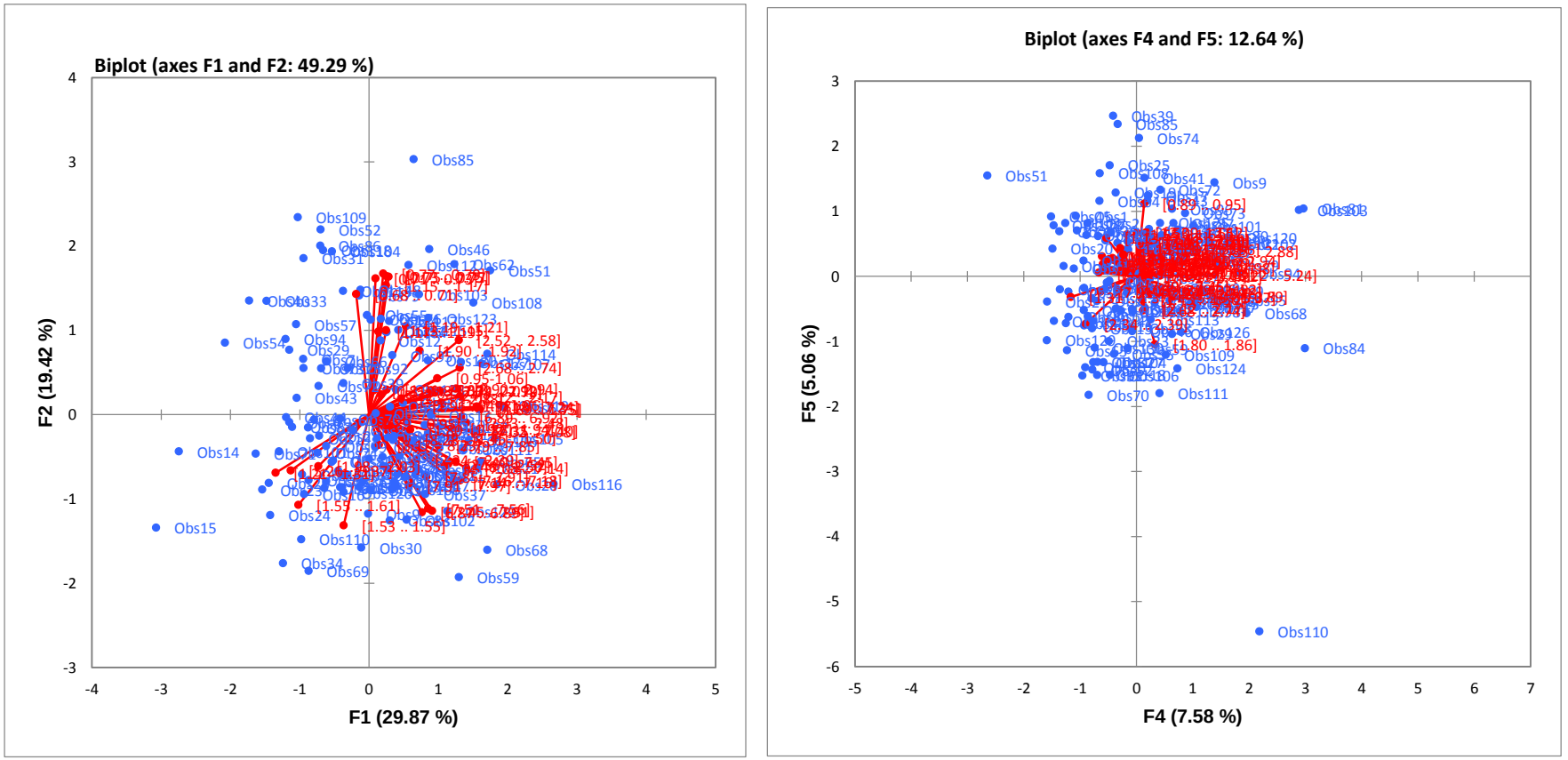

Figure 8. Biplots showing two separate classes in the NPC1 disease dataset, with red representing the NPC1 disease patients and the blue corresponding to the healthy controls/wild treated (WT), where the observations 15, 51, 85 and 110 are outliers.

- **Consideration**

Using PCA, 55 PCs were determined and only the first 14 of them were used to explain 89% of variability of the present NPC1 disease dataset based on these new features, the PCs. Knowing that this requirement was achieved in terms of selecting the necessary PCs, the multivariate logistic regression was applied to complete the principal components regression (PCR) analysis.

- **PCR results related to the NPC1 disease diagnosis – Human model**

- **PCR results based on the tri-ranking techniques (TRTs) defined for potential Biomarker discovery in the NPC1 disease dataset**

  - **PCR modelling results – 90% variability**

NPC1 Disease Diagnosis for the 90% Variability				
Model Statistics	***Coefficient***	***SE***	***t-statistics***	***P Value***
Deviance of the Fit:	67.9777	Null	Null	Null
Degrees of Freedom:	114	Null	Null	Null
Estimated dispersion parameter:	0.832	Null	Null	Null
Feature Number				
0	0.299			>0.05
1	**-1.392**	**0.4191**	**-3.3218**	**0.0009**
2	**-1.2947**	**0.3795**	**-3.4117**	**0.0006**
3	-0.3291			>0.05
4	**-2.7617**	**0.5821**	**-4.7445**	**0**
5	0.0309			>0.05
6	0.3156			>0.05
7	**0.6039**	**0.3059**	**1.9746**	**0.0483**
8	0.1099			>0.05
9	**0.8591**	**0.3379**	**2.5425**	**0.011**
10	-0.6418			>0.05
11	0.7331			>0.05
12	-0.8776			>0.05
13	**1.0659**	**0.3842**	**2.7747**	**0.0055**
14	**0.8635**	**0.4336**	**1.9913**	**0.0464**
15	-0.8216			>0.05
* Feature number 0 is the intercept.	Null	Null	Null	Null
Null	Null	Null	Null	Null

Table 14. 90% variability model and the different statistics, giving PCs coefficient of regression estimate dispersion and deviance of the fit. The lower the coefficients of deviance of the fit and estimate dispersion parameter, the better the model could explain the level of variability in the dataset. Therefore, low value of coefficients of deviance of the fit = 67.9777 and low value of Estimated dispersion parameter = 0.832 correspond to high variability 90% variability with higher overall ranking performance. The features coefficients allow performance of the ranking established by the PCR, and also the different ranking techniques. The very important features

are in this order of importance 13, 14, 9, and 7; based on their coefficient of regression, and a small p value < 0.05. The features with p value > 0,05 are statistically insignificant.

The PCR model uses the equation 42 below (established in CH4) to calculate the contribution of each feature to the relevant principal component, while equation 43 corresponds to the regression line's one:

$$K_{ij} = \sum_{j=1}^{j\leq m} \sum_{i=1}^{i\leq n} r_j \, \alpha_{ji} \qquad : (\mathbf{42})$$

$$Y = \ln\left(\frac{p}{1-p}\right) = \sum_{j=1}^{j\leq m} \sum_{i=1}^{i\leq n} r_j \, \alpha_{ji} \, X_i + Y_o \qquad : (\mathbf{43})$$

r_j *is the coefficient of regression and* α_{ji} *is the coefficient of correlation between and* PCs, *with* m *maximum number of* PCs *considered to explain the required variability level,* n *the maximum number of features in the dataset. The regression is expressed as a probability, The value* p *representing the probability of the individual having the disease. The term* K_{ij} *is the sum of product corresponds to the slot of the regression line. Its value represents the multiplicative factor by which each feature contribute to the regression line* Y.

An example of the calculation of the contribution factor K_{ij} is given for the 60% to make it easier for the reader in term of the calculation length (see Table 22 below). Thus, the PCs coefficients of regression in Table 12 and Table 21 are used to perform the ranking obtained by the PCR and other ranking techniques separately.

- **Chemical shifts/buckets of biomolecules related to NPC1 disease dataset**

Features Ranking	**Chemical shift**	**Probable Biomarkers**
1	[1.55…1.61]	Isocaproate-CH_2
2	**[1.53…1.55]**	**Adipate/n-butyrate**
3	[2.34...2.39]	Pyruvate-CH_3; Glutamate-C_3-CH_2
4	[1.31…1.37]	Lactate-CH_3; 2-Hydroxyisobutyrate-CH_3's
5	[1.21…1.31]	3-D-Hydroxybutyrate-CH_3; L-Fucose-CH_3; and 3-Hydroxyisovalerate-CH_3
6	[7.85...7.91]	1-Methylhistidine-C_4-CH
7	[7.91…7.97]	Quinolinate-C_5-CH
8	[7.45…7.51]	Indoxyl sulphate-C_6-CH and Benzoate-C_3/C_5-CH
9	[1.98…2.03]	2-Hydroxyglutarate-C3-CHB
10	[1.13…1.15]	2,3-Butanediol
11	[5.26...5.37]	α-Glucose
12	[8.28...8.31]	Hypoxanthine-C_2/C_7-CH
13	[0.89…0.95]	Isoleucine-CH_3
14	[1.92…1.94]	Acetate-CH_3
15	[8.17…8.23]	Trigonelline-C_5-CH
16	[0.68…0.71]	Cholesterol-C_{18}-CH_3
17	[7.14...7.16]	Hydroxyphenyl-acetate aromatic ring protons
18	[6.87…6.89]	3-(3-hydroxyphenyl)-3-hydroxypropanoate-C1/C6-CH
19	[7.08…7.14]	Histidine
20	[2.48…2.50]	Glutamine-C_4-CH_2
21	[2.50...2.52]	Methylamine
22	[2.03…2.09]	N-acetylaspartate-CH_3; Nacetylneuraminate-CH_3; and N-acetyl-X-CH_3
23	[7.40…7.45]	Indoxyl-sulphate-C6/C9-CH
24	**[7.51…7.56]**	**Hippurate-C3/C5-CH**

Table 15. Shows the partial chemical shifts buckets related to the NPC1 disease dataset (human model), used in the identification of the different buckets involved in the ranking established by the different techniques employed in this research work.

- **PCR results based on the tri-ranking techniques (TRTs) defined for the potential biomarkers discovery in the NPC1 disease diagnosis (90% variability)**

The percentage of selections are noted correct ranking = CR and incorrect ranking = IR in comments column.

Ranking Method	**Potential Biomarkers (90% Var.)**	**Comments**
Heuristic Method	[1.55...1.61] [1.31…1.37] [2.34...2.39] [1.21...1.31] [1.53…1.55] [7.79...7.85] [7.85...7.91] [5.26...5.37] [8.28...8.31] [7.91..7.97] [7.51…7.56] [1.98...2.03] [7.45 .. 7.51] [1.13...1.15] [1.92...1.94] [0.68...0.71] [7.14…7.16] [6.87…6.89] [8.17...8.23] [2.03...2.09] [7.08…7.14] [2.48...2.50] [7.40…7.45] [0.71…0.73]	Compared to SPC CR = 22/24 = 91.67% IR = 2/24 = 08.33% In blue the 10 most effective biomarkers detected by the heuristic ranking technique – In red the ones incorrectly ranked
The Sum of the Product of the Coefficients (SPC) Method	[1.55...1.61] [1.53…1.55] [2.34…2.39] [1.31…1.37] [1.21…1.31] [7.85…7.91] [7.91…7.97] [7.79...7.85] [1.98...2.03] [7.51...7.56] [7.45…7.51] [1.13…1.15] [5.26...5.37] [8.28...8.31] [0.89…0.95] [1.92…1.94] [8.17…8.23] [0.68...0.71] [7.14...7.16] [6.87…6.89] [7.08…7.14] [2.48…2.50] [2.50...2.52] [2.03…2.09]	Reference Patterns In blue the 10 most effective biomarkers detected by the SPC ranking techniques
The Exponential of the Sum of the Product of the Coefficients (ESPC) Method	[1.55…1.61] [1.53…1.55] [2.34…2.39] [1.31…1.37] [1.21…1.31] [7.85…7.91] [7.91…7.97] [7.79…85] [1.98…2.03] [7.51...7.56] [7.45…7.51] [1.13…1.15] [5.26…5.37] [8.28...8.31] [0.89...0.95] [1.92…1.94] [8.17...8.23] [0.68...0.71] [7.14…7.16] [6.87... 6.89] [7.08...7.14] [2.48…2.50] [2.50...2.52] [2.03...2.09]	Correct Classification Compared to SPC CR = 24/24 = 100% IR = 0/24 = 0% In blue the 10 most effective biomarkers detected by the ESPC ranking techniques

Table 16. 24 most important buckets corresponding to the 24 best markers detected by the tri-ranking techniques and potential biomarkers for the 90% total variability dataset. The blue colour shows the 8 best buckets and the red shows some differences in the feature ranking between the different techniques with regard to the 10 best buckets.

The different PCs and their corresponding r values are obtained by repeating the MLR operation several times. Table 16 above gives the main PCs involved in the 90%, variability that are being analysed. The deviance of the model is 67.98 and the estimated dispersion is < 1. The statistics show that this model successfully fit the dataset under scrutiny, although some argue that the deviance is not a measure of fit in binary regression (Ronneberg, 2015). Nonetheless, it should be noted that the linear combination of the values of r and α permits a final classification of the features with regard to their importance in biomarker discovery regarding the NPC1 disease diagnosis (human model). The result of the NPC1 disease dataset analysis (Table 15) shows that the heuristic and the SPC strategies give almost the same ranking, while potential biomarkers selected by the SVA are also detected by all tri-ranking techniques. However, heuristic is the only one ranking technique which performed differently from the other two models with a value of correct ranking of 91.67%.

- **PCR results based on the tri-ranking techniques (TRTs) employed for potential biomarker discoveries in the NPC1 disease dataset**

  - **PCR modelling results – 80% variability**

NPC1 Disease Diagnosis for the 80% Variability				
Model Statistics	***Coefficient***	***SE***	***t-statistics***	***p value***
Deviance of the Fit:	93.9176	Null	Null	Null
Degrees of Freedom:	120	Null	Null	Null
Estimated dispersion parameter:	1.273	Null	Null	Null
Feature Number				
0	0.3181			>0.05
1	**-0.9565**	0.2982	-3.2077	0.0013
2	**-0.921**	0.2892	-3.1842	0.0015
3	-0.2677			>0.05
4	**-2.105**	0.3833	-5.4918	0

5	0.0892	-	-	>0.05
6	0.28	-	-	>0.05
7	0.4893	-	-	>0.05
8	0.3419	-	-	>0.05
9	**0.6164**	0.2776	2.2201	0.0264
** Feature number 0 is the intercept.	Null	Null	Null	Null
Null	Null	Null	Null	Null

Table 17. 80% variability model and the different statistics, giving PCs coefficient of regression estimate dispersion and deviance of the fit. The relatively low coefficient of deviance and high coefficient of dispersion explained 80% of the total variation and a higher overall ability in ranking performance was achieved. Features 1, 2, 4, and 9 are statistically very important in this feature ranking process, with p value < 0.05. This Table provides different PCs feature coefficients of regression, which are used to perform the following ranking based on the PCR model, and the rankings arising from use of the different techniques. Features with p value > 0.05 are not statistically significant, therefore their standard error (SE) and t-statistic values are not provided. The feature 0 is the intercept, and the ranking results are obtained via the approach outlined below. Statistical significant features with p value < 0.05 are in bold.

- **PCR Results Based on the tri-ranking techniques (TRTs) defined for Potential Biomarker Discoveries in the NPC1 Disease Diagnosis (80% Variability)**

Ranking Method	**Potential Biomarkers (80% Var.)**	**Comment**
Heuristic Method	[2.34…2.39] [1.31…1.37] [1.55…1.61] [7.91…7.97] [7.85…7.91] [1.21…1.31] [7.51…7.56] [1.53…1.55] [5.26…5.37] [7.79…7.85] [1.92…1.94] [8.17…8.23] [8.28…8.31] [7.45…7.51] [2.03…2.09] [1.98…2.03] [0.68…0.71] [1.13…1.15] [6.87…6.89] [7.08…7.14] [7.14…7.16] [2.50…2.52] [2.48…2.50] [7.40…7.45]	Compared to SPC CR = 23/24 = 95.83 % IR = 1/24 = 4.17% In blue the 10 most effective biomarkers detected by the heuristic ranking technique
The Sum of the Product of the Coefficients (SPC) Method	[1.55…1.61] [2.34…2.39] [1.31…1.37] [1.53…1.55] [8.17…8.23] [1.21...1.31] [7.85…7.91] [7.91…7.97] [1.98…2.03] [7.51…7.56] [7.45…7.51] [1.92…1.94]	The Pattern detected by SPC are used as Reference Patterns

	[7.79…7.85] [0.89…0.95] [8.28…8.31] [5.26…5.37] [2.03…2.09] [1.13…1.15] [0.68…0.71] [2.50…2.52] [6.87…6.89] [7.08…7.14] [7.14…7.16] [2.48…2.50]	In blue the 10 most effective biomarkers detected by the SPC ranking techniques
The Exponential of the Sum of the Product of the Coefficients (ESPC) Method	[1.55…1.61] [2.34…2.39] [1.31…1.37] [1.53…1.55] [8.17…8.23] [1.21…1.31] [7.85…7.91] [7.91…7.97] [1.98…2.03] [7.51…7.56] [7.45…7.51] [1.92…1.94] [7.79…7.85] [0.89…0.95] [8.28…8.31] [5.26…5.37] [2.03…2.09] [1.13…1.15] [0.68…0.71] [2.50…2.52] [6.87…6.89] [7.08…7.14] [7.14…7.16] [2.48…2.50]	Correct Classification Compared to SPC CR = 24/24 =100 % IR = 0/24 = 0% In blue the 10 most effective biomarkers detected by the ESPC ranking techniques

Table 18. 24 chemical most important shifts corresponding to the 24 best markers detected by the tri-ranking techniques and the potential biomarkers for the model with a total 80% variability. The blue colour shows the 10 most important chemical shifts/"biomarkers" and the red one shows some differences in the features ranking between the different techniques.

Table 18 above gives the main PCs involved in the of 80% maximum variability model that is analysed in this section. The deviance of the model is = 93.918, and the estimated dispersion is = 1.273. The statistics show that the current model still effectively fits the data under scrutiny, but to a lesser degree than the previous with case 90% variability. The three ranking models have detected and ranked amongst the 10 most effective features, 8 most effective biomarkers. In addition, 23 out of 24 main biomarkers are detected by the three different ranking techniques. The heuristic model is the only one performing differently, a value of CR = 95.83% being obtained in this respect is important. Therefore, the difference observed is probably related to the fact that the heuristic technique does involve a rapid, although not a perfect solution.

- **PCR results based on the tri-ranking techniques (TRTs) selected for potential biomarkers discovery in the NPC1 disease dataset**

  - **PCR Modelling Results – 70% Variability**

NPC1 Disease Diagnosis for the 70% Variability				
Model Statistics	***Coefficient***	***SE***	***t-statistics***	***P Value***
Deviance of the Fit:	104.71	Null	Null	Null
Degrees of Freedom:	124	Null	Null	Null
Estimated dispersion parameter:	1.397	Null	Null	Null
Feature Number				
0	0.2509			>0.05
1	**-0.9945**	**0.2893**	**-3.4376**	**0.0006**
2	**-0.9352**	**0.275**	**-3.4006**	**0.0007**
3	-0.246			>0.05
4	**-2.0131**	**0.3615**	**-5.5685**	**0**
5	0.0537			>0.05
* Feature number 0 is the intercept.	Null	Null	Null	Null
Null	Null	Null	Null	Null

Table 19. 70% variability model and the different statistics, giving PCs coefficient of regression estimate dispersion and deviance of the fit. The high coefficient of deviance of the fit = 104.7 and the estimate dispersion parameter = 1.397 (> 1) could explain the lower overall ranking performance achieved. The features with p values > 0.05 are not statistically significant, and hence the standard error (SE) and the t-statistic values are not provided for these. The feature "o" is the intercept and the ranking results obtained are provided below based on PCR model and the differing features coefficients of regression given in Table 19.

- **PCR results based on the tri-ranking techniques (TRTs) selected for potential biomarkers discovery in the NPC1 disease diagnosis dataset (70% Variability)**

Ranking Method	**Potential Biomarkers (70% Var.)**	**Comment**
Heuristic Method	[1.31…1.37] [2.34…2.39] [7.91…7.97] [1.55…1.61] [1.21…1.31] [7.51…7.56] [5.26…5.37] [7.85…7.91] [7.79…7.85] [1.53…1.55] [7.45…7.51] [8.28…8.31]	Compared to SPC CR =21/24 = 87.5% IR =3/24 =12.5 %

	[0.68…0.71] [1.92…1.94] [8.17…8.23] [1.98…2.03] [7.08…7.14] [7.14…7.16] [6.87…6.89] [1.13…1.15] [2.03…2.09] [0.71…0.73] [7.16…7.18] [2.48…2.50]	. In blue the same 10 most effective biomarkers detected by the heuristic ranking technique like the ones detected by SPC.
The Sum of the Product of the Coefficients (SPC) Method	[1.55…1.61] [1.31…1.37] [2.34…2.39] [1.53…1.55] [1.21…1.31] [7.91…7.97] [8.17…8.23] [7.51…7.56] [8.28…8.31] [7.85…7.91] [1.98…2.03] [7.79…7.85] [7.45…7.51] [5.26…5.37] [0.89…0.95] [1.92…1.94] [0.68…0.71] [7.14…7.16] [6.87…6.89] [2.03…2.09] [1.13…1.15] [7.08…7.14] [7.40…7.45] [2.50…2.5]	. The Pattern detected by SPC are used as Reference Patterns. .In blue the 10 most effective biomarkers detected by the SPC ranking techniques.
The Exponential of the Sum of the Product of the Coefficients (ESPC) Method	[1.55…1.61] [1.31…1.37] [2.34…2.39] [1.53…1.55] [1.21…1.31] [7.91…7.97] [8.17…8.23] [7.51…7.56] [8.28…8.31] [7.85…7.91] [1.98…2.03] [7.79…7.85] [7.45…7.51] [5.26…5.37] [0.89…0.95] [1.92…1.94] [0.68…0.71] [7.14…7.16] [6.87…6.89] [2.03…2.09] [1.13…1.15] [7.08…7.14] [7.40…7.45] [2.50…2.52]	Correct Classification Compared to SPC CR = 24/24 = 100% IR =0/24 = 0% . In blue the same 10 most effective biomarkers detected by the ESPC ranking technique like the ones detected by SPC.

Table 20. 24 most important chemical shifts corresponding to the 24 most prominent biomarkers detected by the tri-ranking techniques and the potential biomarkers for the 70% variability. The blue colour showing the same 10 best chemical shift biomarker buckets detected by all the tri-ranking techniques, including heuristic, SPC, and ESPC. Based on the reference patterns these correspond to the following molecules presented below: isocaproate-CH_2; actate-CH_3 2-hydroxyisobutyrate-CH_3; pyruvate-CH_3, glutamate-C_3-CH_2; adipate/n-butyrate, 3-hydroxybutyrate-CH_3, L-fucose-CH_3, 3-hydroxyisovalerate-CH_3; quinolinate-C_5-

CH; hippurate-C3/C5-CH; hypoxanthine-C_2/C_7-CH; and 1-methylhistidine-C_4-CH. The red shows some difference in features rankings between the heuristics ranking technique, the SPC, and the ESPC.

Table 20 above gives the main PCs involved in the 70%, variability being analysed in this section. The deviance of the model is = 104.71, and the estimated dispersion is = 1.397. The statistics show that the current model still fit the data being analysed, but to a lesser degree than the model involving 90% and 80% total variability. The three ranking models commonly detected the same 10 main biomarker buckets. In addition, 21 out of 24 main buckets are detected by the three different ranking techniques. The heuristic model was the only one performing differently, with a CR value of 87.5%. Additionally, this ranking technique gives preference to speed at the expenses of both precision and accuracy, consequently does not detect isoleucine (0.89-0.95 ppm) amongst the 24 main biomarkers.

- **PCR results based on the tri-ranking techniques (TRTs) selected for potential biomarkers discovery in the NPC1 Disease Dataset**
  - **PCR Modelling Results – 60% Variability**

NPC1 Disease Diagnosis for the 60% Variability				
Model Statistics	***Coefficient***	***SE***	***t-statistics***	***p value***
Deviance of the Fit:	104.7645	Null	Null	Null
Degrees of Freedom:	125	Null	Null	Null
Estimated dispersion parameter:	1.39	Null	Null	Null
Feature Number				
0	**0.248**	-	-	>0.05
1	**-0.9952**	0.289	-3.4442	0.0006
2	**-0.9388**	0.2753	-3.4107	0.0006
3	**-0.2453**	-	-	>0.05
4	**-2.0141**	0.3611	-5.5775	0
** Feature number 0 is the intercept.	Null	Null	Null	Null
Null	Null	Null	Null	Null

Table 21. Ranking of PCs based on the regression coefficient and a percentage of total variability of 60%. The high coefficient of deviance of the fit = 104.7645 and the estimate

dispersion parameter = 1.39 i.e. (> 1) could explain the lower overall ranking performance achieved in this model. For the features that are not statistically significant (p value > 0.05), the standard error (SE) and the t-statistic values are not provided and the ranking results obtained are as follow. An example of the calculation for 60% variability for the sum product of the coefficients is: $K_{ij} = \sum \mathbf{r\alpha} = \alpha_1 r_1 + \alpha_2 r_2 + \alpha_3 r_3 + \alpha_4 r_4$: (**42**)

Features	α_1	r_1	α_2	r_2	α_3	r_3	α_4	r_4	$\sum \mathbf{r\alpha}$
[1.55…1.61] Isocaproate-CH_2	-0.575	-0.9952	**-0.604**	-0.9388	-0.088	-0.2453	-0.095	-2.0141	1.352078
[0.89…0.95] Isoleucine-CH_3	-0.087	-0.9952	-0.056	-0.9388	**-0.610**	-0.2453	0.078	-2.0141	0.131828

Table 42. Example of calculation related to the SPC ranking for the 60% maximum variability model for the NPC1 disease diagnosis dataset. The example is showing how the following biomarkers were ranked using the values of $K_{ij} = \sum \mathbf{r\alpha}$, including isocoprate-$CH_2$, and isoleucine-$CH_3$. The calculation is applied to the feature ranking in relation to the different variability, and the results obtained for the 60% maximum variance model is provided below. The same calculation technique is applied to the ESPC but using the exponential values of the products.

Ranking Method	**Potential Biomarkers (60% Var.)**	**Comment**
Heuristic Method	[1.31…1.37] [2.34…2.39] [7.91...7.97] [1.55…1.61] [1.21…1.31] [7.51...7.56] [5.26…5.37] [7.85…7.91] [7.79…7.85] [1.53…1.55] [7.45…7.51] [8.28…8.31] [0.68…0.71] [8.17…8.23] [1.98…2.03] [7.08…7.14] [7.14…7.16] [1.92…1.94] [6.87…6.89] [1.13…1.15] [2.03…2.09] [0.71…0.73] [7.16…7.18] [2.48…2.50]	Compared to SPC CR = 21/24 = 87.5% IR = 3/24 = 12.5% In blue the same 10 most effective biomarkers detected by the heuristic ranking technique like the ones detected by SPC.
The Sum of the Product of the	[1.55…1.61] [1.31…1.37] [2.34...2.39] [1.53...1.55] [1.21…1.31] [7.91...7.97]	The Pattern detected by SPC are used as Reference

Coefficients (SPC) Method	[8.17...8.23] [7.51…7.56] [8.28...8.31] [7.85…7.91] [1.98…2.03] [7.79…7.85] [7.45…7.51] [5.26…5.37] [0.89...0.95] [1.92…1.94] [0.68…0.71] [7.14…7.16] [6.87…6.89] [2.03…2.09] [7.08…7.14] [1.13…1.15] [7.40…7.45] [2.50…2.52]	Patterns. In blue the 10 most effective biomarkers detected by the SPC ranking techniques.
The Exponential of the Sum of the Product of the Coefficients (ESPC) Method	[1.55…1.61] [1.31…1.37] [2.34…2.39] [1.53...1.55] [1.21…1.31] [7.91…7.97] [8.17…8.23] [7.51...7.56] [8.28...8.31] [7.85...7.91] [1.98…2.03] [7.79…7.85] [7.45…7.51] [5.26…5.37] [0.89...0.95] [1.92...1.94] [0.68…0.71] [7.14…7.16] [6.87…6.89] [2.03...2.09] [7.08…7.14] [1.13…1.15] [7.40…7.45] [2.50...2.52]	Correct Classification Compared to SPC CR = 24/24 = 100% IR = 0/24 = 0% In blue the same 10 most effective biomarkers detected by the ESPC ranking technique like the ones detected by SPC.

Table 23. Display of potential biomarkers based on the three main techniques developed for the 60% total variance model. The blue colour shows the 10 main chemical shift buckets and the red ones the main differences in feature ranking obtained between the heuristic and the ESPC techniques.

The deviance of the model is 104.7645, and the estimated dispersion is 1.39 (Table 21). The statistics show that the current model still fits the data being analysed effectively, but to a lesser degree than those observed with 90 and 80% variability. The three models commonly detected the same 10 main biomarkers. In addition, 21 out of 24 major biomarkers are detected by the three different ranking techniques (Table 23). The heuristic model was the only one which performed differently, and its correct ranking value of 87.5% enables an acceptable comparison between the techniques. However, this technique did not detect isoleucine (0.89-0.95 ppm) in view of its propensity to favour speed at the expense of precision and accuracy.

A comparison between the four variability levels is next presented, with special regard to the SPC approaches utilised as ranking reference models in view of their higher performance level.

- **Comparing the Four Types of Variability**

90% Variability	**80% Variability**
[1.55...1.61] [1.53…1.55] [2.34…2.39] [1.31…1.37] [1.21…1.31] [7.85…7.91] [7.91…7.97] [7.79...7.85] [1.98...2.03] [7.51...7.56] [7.45…7.51] [1.13…1.15] [5.26...5.37] [8.28...8.31] [0.89…0.95] [1.92…1.94] [8.17…8.23] [0.68…0.71] [7.14...7.16] [6.87…6.89] [7.08…7.14] [2.48…2.50] [2.50...2.52] [2.03…2.09]	[1.55…1.61] [2.34…2.39] [1.31... .37] [1.53…1.55] [8.17...8.23] [1.21..1.31] [7.85…7.91] [7.91...7.97] [1.98…2.03] [7.51…7.56] [7.45…7.51] [1.92…1.94] [7.79…7.85] [0.89...0.95] [8.28…8.31] [5.26…5.37] [2.03…2.09] [1.13...1.15] [0.68…0.71] [2.50...2.52] [6.87…6.89] [7.08...7.14] [7.14…7.16] [2.48...2.50]
70% Variability	**60% Variability**
[1.55…1.61] [1.31...1.37] [2.34…2.39] [1.53…1.55] [1.21…1.31] [7.91...7.97] [8.17...8.23] [7.51…7.56] [8.28…8.31] [7.85…7.91] [1.98…2.03] [7.79…7.85] [7.45…7.51] [5.26…5.37] [0.89…0.95] [1.92…1.94] [0.68…0.71] [7.14…7.16] [6.87…6.89] [2.03…2.09] [1.13…1.15] [7.08…7.14] [7.40…7.45] [2.50…2.52]	[1.55 .. 1.61] [1.31 .. 1.37] [2.34…2.39] [1.53…1.55] [1.21…1.31] [7.91…7.97] [8.17…8.23] [7.51…7.56] [8.28...8.31] [7.85…7.91] [1.98…2.03] [7.79…7.85] [7.45…7.51] [5.26…5.37] [0.89...0.95] [1.92…1.94] [0.68…0.71] [7.14…7.16] [6.87…6.89] [2.03…2.09] [7.08...7.14] [1.13…1.15] [7.40…7.45] [2.50…2.52]

Table 24. Sequence of the ranking of potential biomarkers (as ^{1}H NMR chemical shift buckets) detected. The consistency and the consecutiveness of the different biomarkers discovered are shown in green, while the ones in blue does not follow this sequence of the 10 best ranked biomolecules. Isocaproate, adipate/n-butyrate, pyruvate, glutamate, and lactate represent some of the major biomarkers highly ranked by all four variability levels. Thus, those depicted in green are correctly ranked in the top 10 biomarkers amongst the 24, while those in blue are incorrectly ranked amongst these top 10. The ones in black have been detected by the four variabilities levels amongst the 24, while those in red have not been detected by more than 2 levels of variability. These are differing common markers amongst the 24 best potential biomarkers. Additionally, the ranking of the top 10 best potential biomarkers show some consistency for 90% (2 biomarkers are misclassified) and the rest 80, 70 and 60% with 3 biomarkers misclassified. However, the ones in red have not been detected by all four levels of variability.

The consistency, consecutiveness and importance of the biomarkers discovered revealed that 80% and 90% total variability are models to be considered as the best and second best available for NPC1 disease diagnosis in blood plasma samples. However, more statistics can be gathered to ascertain this classification model. Comparison of these models based on the percentage of variability (below) assists in this respect.

- **Comparisons of models and ranking techniques (TRTs) based on variability level**

Level of Variability (%)	SPC Method (%)	Heuristic Method with the CR (%)	Model Deviance of the Fit	Parameters Estimated Dispersion	Degrees of Freedom (DoF)
90	100	91.67	67.9777	0.832	114
80	100	95.83	93.9176	1.273	120
70	100	87.50	104.71	1.397	124
60	100	87.50	104.7645	1.39	125

Table 25. Percentage (%) of the correct ranking (CR) perform by the Heuristic and SPC (identical to ESPC) strategies. Also shown are performance statistics related to each level of variability. The smaller the model deviance of the fit, and the parameters estimated, the more reliable are the variability model performances. 80% and 90% appear to be the best model in terms of their correct ranking ability.

- **Considerations**

1. Heuristic model provides the only ranking technique with different performances, and therefore its level of correct ranking (CR) will be determinant for comparisons between the different models, or those between different levels of variability.

2. SPC and ESPC ranking methods achieve the same high level of correct rankings or selection, and are therefore the best ranking techniques based on these criteria.

3. Ranking abilities confirmed that 80% and then 90% respectively were the best and second-best variability percentages to be considered for the diagnosis potential of the NPC1 disease.

It should be noted that 80% variability, although having a high level of deviance of the fit (93.9176 > 67.9777), and also a high level of parameters estimated, i.e. dispersion (1.273 > 0.832) than 90% variability, performed better in terms of correct ranking than the second model mentioned. Indeed, it is expected that the smaller the deviance of the model fit and parameter of estimate dispersion, the more improved the level of variability and the model. Finally, the models' ability for correctly ranking the features, i.e. 95.83% CR for 80% model against 91.67% CR for 90% model, allows us to choose 80% model to be considered as the best level of variability followed by 90% as second best for this dataset.

The Table below gives the full range of 24 major potential biomarkers, together with those discovered by the PCR model applied in this study. Comparisons are made with already established biomarkers from literature (Ruiz-Rodado et al., 2014; Ruiz-Rodado, 2016).

- **Molecular assignment of the 24 main biomarkers in the NPC1 disease diagnosis**

Features Ranking	Chemical shift	Probable Biomarkers	Comment on the Biomarkers Discovered
1	[1.55 .. 1.61]	Isocaproate-CH_2	Potential Biomarker
2	[1.53 .. 1.55]	***Adipate/n-butyrate***	New Biomarkers
3	[2.34 .. 2.39]	Pyruvate-CH_3; Glutamate-C_3-CH_2	Existing Biomarkers
4	[1.31 .. 1.37]	Lactate-CH_3; 2-Hydroxyisobutyrate-CH_3's	Potential Biomarkers
5	[1.21 - 1.31]	3-D-Hydroxybutyrate-CH_3; L-Fucose-CH_3; and 3-Hydroxyisovalerate-CH_3	Potential Biomarkers
6	[7.85 .. 7.91]	1-Methylhistidine-C_4-CH	Potential Biomarker
7	[7.91 .. 7.97]	Quinolinate-C_5-CH	Potential Biomarker
8	[7.45…7.51]	Indoxyl sulphate-C_6-CH and Benzoate-C_3/C_5-CH	Potential Biomarkers

	9	[1.98…2.03]	2-Hydroxyglutarate-C3-CHB	Potential Biomarker
	10	[1.13…1.15]	2,3-Butanediol	
	11	[5.26...5.37]	α-Glucose	Existing Biomarker
…	12	[8.28...8.31]	Hypoxanthine-C_2/C_7-CH	Potential Biomarker
	13	**[0.89…0.95]**	Isoleucine-CH_3	Existing Biomarker
	14	[1.92…1.94]	Acetate-CH_3	Existing Biomarker
	15	[8.17…8.23]	Trigonelline-C_5-CH	Potential Biomarker
	16	[0.68 .. 0.71]	Cholesterol-C_{18}-CH_3's	Existing Biomarker
	17	[7.14...7.16]	Hydroxyphenyl-acetate aromatic ring protons	Potential Biomarker
	18	[6.87…6.89]	3-(3-hydroxyphenyl)-3-hydroxypropanoate-C1/C6-CH	Potential Biomarker
	19	[7.08…7.14]	Histidine	Existing Biomarker
	20	**[2.48…2.50]**	Glutamine-C_4-CH_2	Existing Biomarker
	21	[2.50...2.52]	Methylamine	Existing Biomarker
	22	[2.03…2.09]	N-acetylaspartate-CH_3; Nacetylneuraminate-CH_3; and N-acetyl-X-CH_3	Potential Biomarkers
	23	[7.40…7.45]	Indoxyl-sulphate-C6/C9-CH	Potential Biomarker
	24	[7.51…7.56]	***Hippurate*-C3/C5-CH**	New Biomarker

Table 26. Existing, potential, and probable new biomarkers discovered by the PCR model for NPC1 disease diagnosis (human model). Pyruvate (3^{rd}), glutamate (3^{rd}), and isoleucine (13^{th}) were detected amongst the 24 best biomarkers; while adipate/n-butyrate (2^{nd}), and hippurate (24^{th}) appeared as probable new biomarker in the NPC1 disease diagnosis.

Results arising from this analysis using the PCR techniques, i.e. the combination of PCA and MLR, permitted the detection of probable biomarkers for NPC1 disease diagnosis, including adipate/n-butyrate, hydroxyphenyl-acetate, and hippurate.

In this respect, the TRTs have all detected the same main biomarkers overall, However, the top 10 differ with changes in variability. This is principally ascribable to the knowledge that the more rapidly, and heuristic technique displays different results, and is based on finding an

approximate solution, hence favour the speed at the expenses of precision and accuracy. Additionally, the SPC and the ESPC approaches give the same ranking results. Moreover, the 80% variability model was found to be the best level of variability to be incorporated amongst the four explored.

However, the main biomarkers discovered by the tri-ranking techniques are predominantly the same, and include nearly all the biomarkers discovered, by application of the SVA model. Finally, the list of the 7 best biomarkers formally detected by all four variability levels in relation to the PCR model selection is adipate/n-butyrate, isocaproate-CH_2, **pyruvate-CH_3, glutamate-C_3-CH_2, lactate-CH_3, 2-hydroxyisobutyrate-CH_3, 3-hydroxybutyrate-CH_3, L-fucose-CH_3, 3-hydroxyisovalerate-CH_3**, 1-methylhistidine-C_4-CH, and quinolinate-C_5-CH. Amongst these, the main biomarkers suggested by the SVA model are highlighted in bold. These markers detected by the SVA approach are important markers for NPC1 disease diagnosis. Other important biomolecules in the NPC1 disease diagnosis were detected, including glutamate, histidine, etc., and will be discussed further in the next sub-section.

5.3. Biomarker Pathway Analysis in the NPC1 Disease Research

5.3.1. Data Normalisation Results

The pathway analysis was performed using MetaboAnalyst 3.0 version (comprehensive tools for metabolomics data analysis). In order to carry out this pathway analysis, data was firstly pre-processed to ensure that data correspond to the software criteria. No missing value estimates, were performed. All features with more than 50% missing values were removed, whilst the remaining missing values were replaced with the column features mean values. In addition, data filtration was performed in order to improve the analysis results by removing the proportion of tests that are truly null, leaving only those that are differentially expressed. Filtering is used to reduce the number of tests, but it increases the power of detection of true differences (Hackstadt and Hess, 2009). Furthermore, the following transformations were performed, and these included, sample normalisation by the sum, data transformation by cube root transformation, finally data was scaled using Pareto scaling which involves in mean-

centring and division by the square root of the standard deviation of each variable. The normalisation results are shown below.

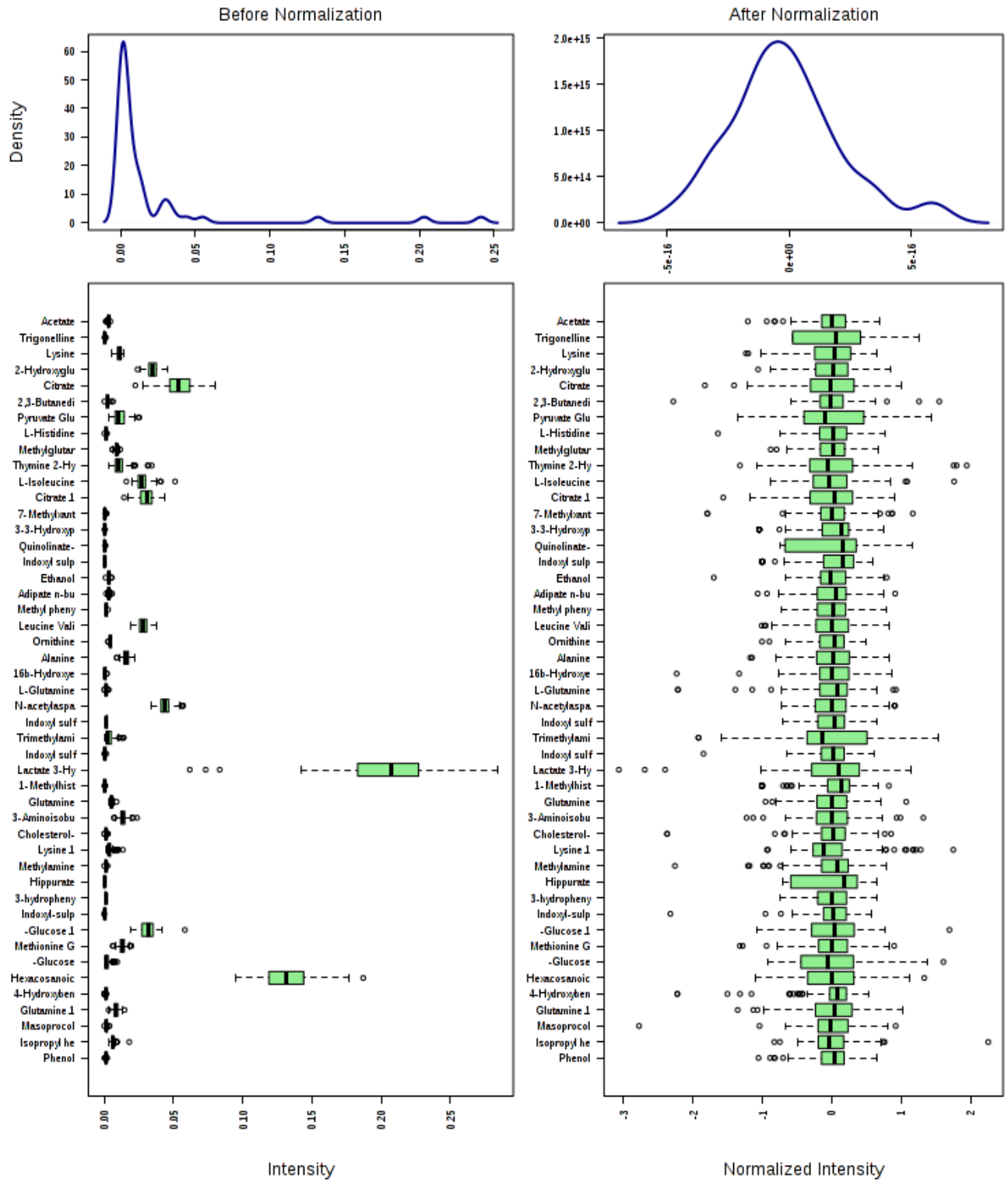

Figure 9. Results of the sample normalisation by the sum, cube-root data transformation and the Pareto data scaling applied on the NPC1 disease dataset. The features density distribution squeeze to the right is after normalisation represented by a normal bell-shaped curve meaning a normal distribution of these features density.

One-way analysis of variance (ANOVA) was performed to found out whether there are statistical difference between the mean values of NPC1 disease features or group of features. A correlation diagram based on Pearson r value with dendrograms applied to the columns and rows highlights results of the relationships between the different metabolites present in the NPC1 disease dataset.

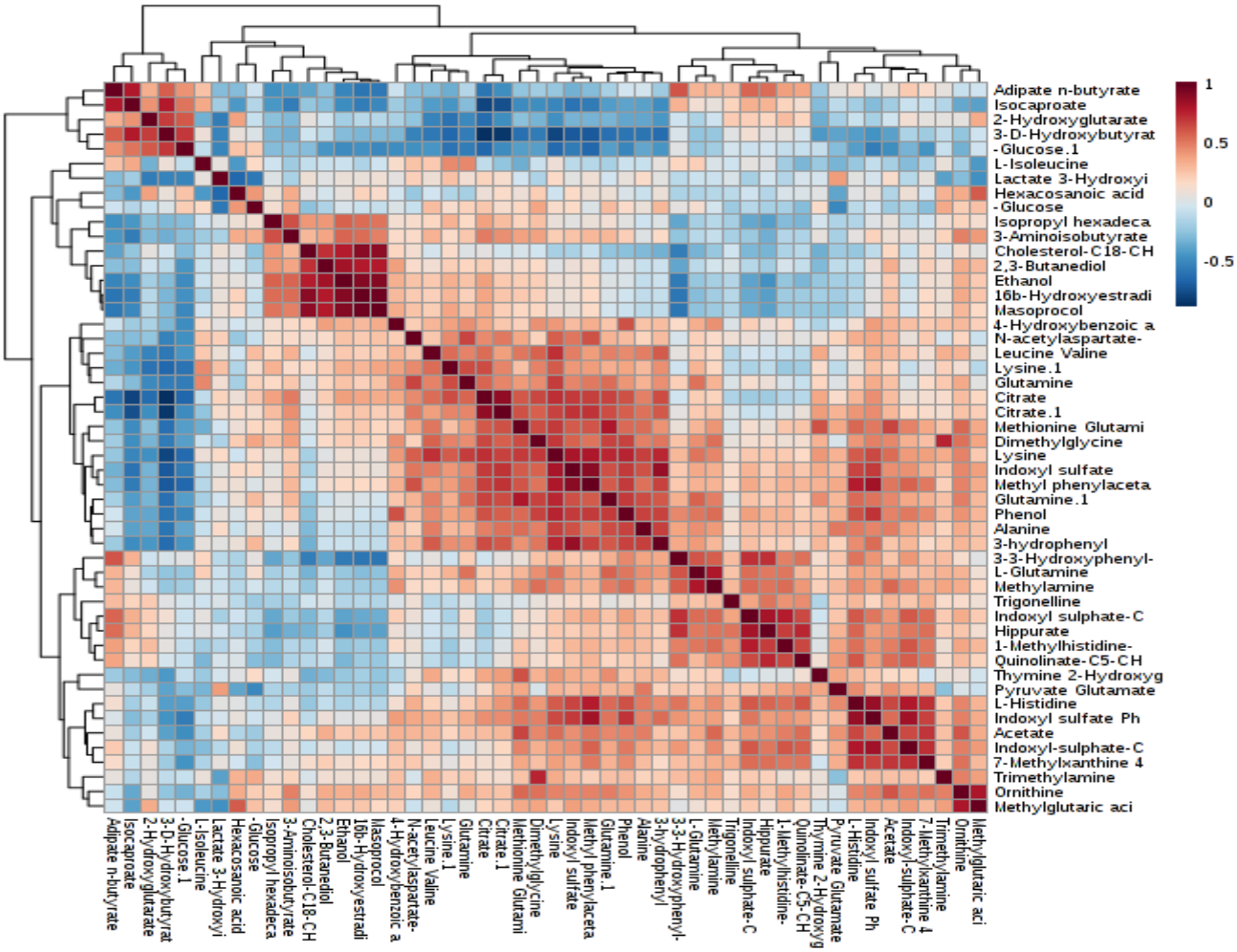

Figure 10. Correlation diagram based on the Pearson's r value with dendrograms applied to the columns and rows, showing relationships between the different potential biomarkers present in the NPC1 disease dataset. Strong correlation areas are shown in brown and dark-brown.

The correlation diagram revealed a strong correlation zone on the y-axis at 4-hydroxybenzoic acid up to the methylglutarate to the same metabolites on the x-axis. This agglomerate hierarchy clustering technique used the importance of biomolecules in the disease development to group them together in clusters. Hence, some very strong correlation zones appear such as the ones between alanine and the leucine, valine, lysine, methionine, Citrate, 3-hydrophnyl, phenol, L-histidine, indoxyl sulfate, glutamine, methyl -phenylacetate, glutamine-1, and the dimethylglycine as shown on the correlation diagram (Figure 10). Indeed, alanine biosynthesis is shown in the pathway analysis presented below.

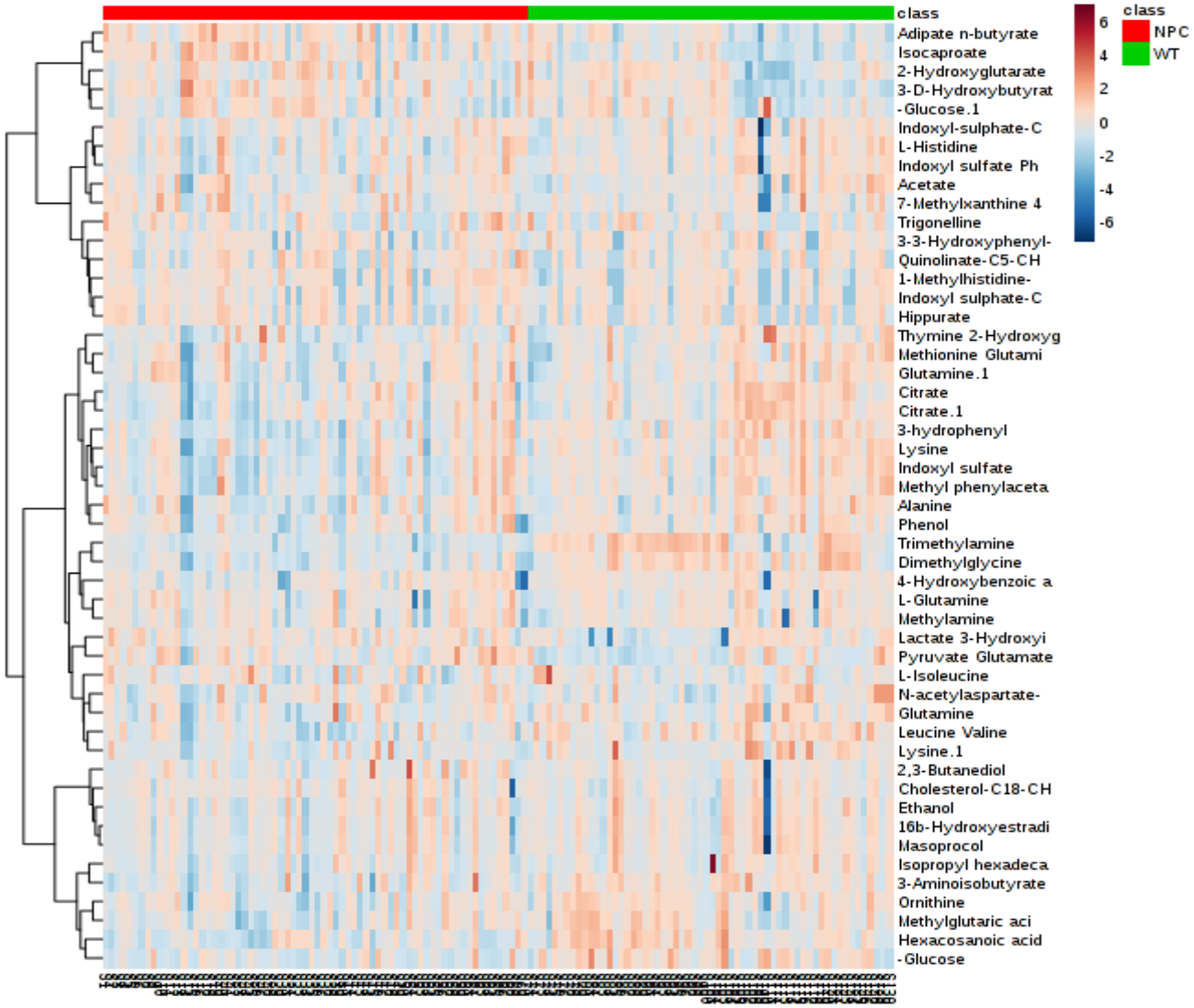

Figure 11. Heatmap using the Euclidean distance presents metabolites using the clustering ward with t-test, with the samples not reorganised which allows to obtain two separate classes as label with NPC in red and WT in blue. The data source was normalised and the autoscaled feature standardisation was also applied. The two main classes showed were the NPC1 disease

group (NPC) and the wild treated (WT) or healthy control group in the plasma dataset. Dendrograms generated using the hierarchal clustering algorithm have been applied in order to aggregate the rows and columns of the heatmap, with the left appearing cooler that is lower presence of NPC1 disease patients or higher presence of healthy control individuals, and the right hotter with a higher presence of NPC1 disease patients in this blood plasma dataset (human-model).

Heatmap shows a high to very high concentration of the 3 top biomarkers including 3-hydroxybutyrate, isocaproate-CH_2, and adipate in the NPC1 disease class when compared to the control group. This is further supported by the light-brown colour and dark-brown colour on the top right corner of the heatmap (Figure 11). However, trimethylamine, dimethylglycine and the methylglutarate are high in the control group. The pathways followed by different metabolites can shed more light on this study.

Pathway Name	Match Status	p	-log(p)	Holm p	FDR	Impact	Details
Aminoacyl-tRNA biosynthesis	10/75	3.1019E-7	14.986	**2.4815E-5**	**2.4815E-5**	0.05634	KEGG
Phenylalanine metabolism	6/45	9.6042E-5	9.2507	**0.0075874**	**0.0038417**	0.15056	KEGG SMP
Glycolysis or Gluconeogenesis	5/31	1.5408E-4	8.778	**0.012018**	**0.0041089**	0.1067	KEGG SMP SMP
Valine, leucine and isoleucine biosynthesis	4/27	0.0010513	6.8577	0.08095	0.021026	0.06148	KEGG SMP
Nitrogen metabolism	4/39	0.0042331	5.4648	0.32172	0.061099	0.0	KEGG SMP
Taurine and hypotaurine metabolism	3/20	0.0045825	5.3855	0.34368	0.061099	0.05395	KEGG SMP
Alanine, aspartate and glutamate metabolism	3/24	0.0077507	4.86	0.57355	0.08858	0.26401	KEGG SMP SMP SMP
D-Glutamine and D-glutamate metabolism	2/11	0.014799	4.2132	1.0	0.14799	0.35294	KEGG SMP
Pyruvate metabolism	3/32	0.017244	4.0603	1.0	0.15328	0.41957	KEGG SMP
Valine, leucine and isoleucine degradation	3/40	0.031199	3.4674	1.0	0.2269	0.02232	KEGG SMP
Butanoate metabolism	3/40	0.031199	3.4674	1.0	0.2269	0.10063	KEGG SMP

Nicotinate and nicotinamide metabolism	3/44	0.039865	3.2223	1.0	0.25529	0.02448	KEGG SMP
Tyrosine metabolism	4/76	0.041485	3.1824	1.0	0.25529	0.04724	KEGG SMP SMP
Citrate cycle (TCA cycle)	2/20	0.046311	3.0724	1.0	0.26464	0.15351	KEGG SMP
Selenoamino acid metabolism	2/22	0.055089	2.8988	1.0	0.29381	0.00321	KEGG SMP
Cysteine and methionine metabolism	3/56	0.072321	2.6266	1.0	0.35235	0.05455	KEGG SMP SMP
Pantothenate and CoA biosynthesis	2/27	0.079279	2.5348	1.0	0.35235	0.0	KEGG SMP
Phenylalanine, tyrosine and tryptophan biosynthesis	2/27	0.079279	2.5348	1.0	0.35235	0.008	KEGG SMP
Pyrimidine metabolism	3/60	0.085146	2.4634	1.0	0.35851	0.04303	KEGG SMP
Synthesis and degradation of ketone bodies	1/6	0.10033	2.2993	1.0	0.40133	0.0	KEGG SMP

Table 27. NPC1 Disease Metabolomic Pathway Analysis, showing 3 main pathways to be significantly important in the disease process. The low Holm adjusted p value and a small value of FDR (< 0.05) in bold show that these pathways are strongly expressed regarding the NPC1 disease aetiology based on a human model.

The information inferred from the Table 27 is used below to analyse the most significant pathways detected in the NCP1 disease dataset. Based on the value of the Holm adjusted p value and the false detection rate FDR the 3 first pathways are shown to be significantly involved in the current NPC1 disease process underlying transformations at the molecular level. The pathway analysis highlights the involvement of the different biomolecules in the disease development, including the transformation of one biomolecule to the next.

5.3.2. Pathway Analysis Results

Metabolomic pathway analysis conducted showed that several transformations appeared significantly involved in NPC1 disease progression. The changes observed included aminoacyl-tRNA biosynthesis, phenylalanine metabolism, and glycolysis or gluconeogenesis. The Holm adjusted p-values of these pathways showed their strong differential expression in

the present disease aetiology. These pathways exhibit a Holm adjusted p value < 0.05 with respectively 2.4815E-5, 0.0075874, and 0.012018. These Holm adjusted p values lower than 0.05 indicate a higher level of statistical significance of the different pathways in the current analysis. The information obtained from Table 27 is further corroborated by the most significant pathways detected and their analysis (Ruiz-Rodado et al., 2014).

- **Overview of Pathway Analysis**

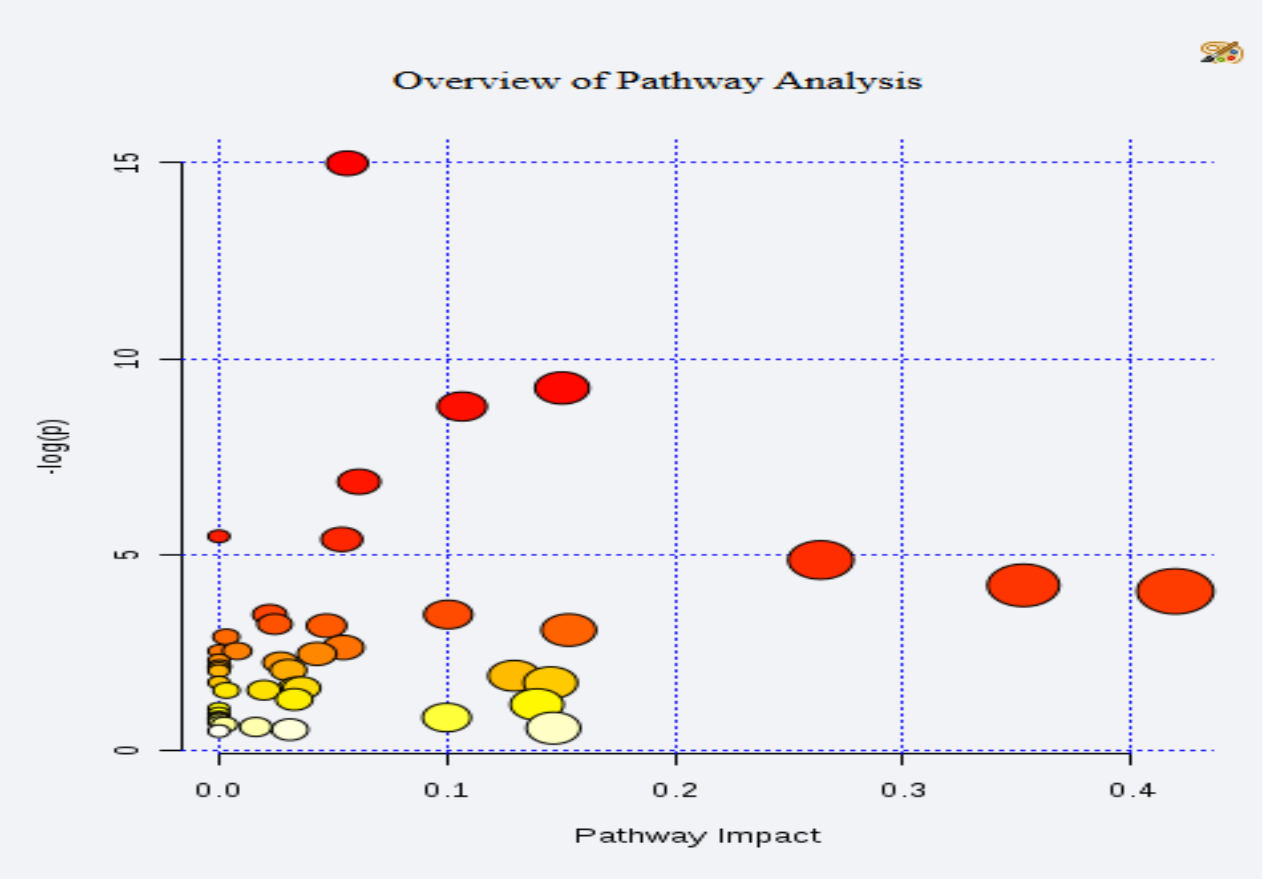

Figure 12. Pathway overview showing from top to bottom the first three main pathways, with their -log(p) value, including aminoacyl-tRNA biosynthesis (14.986), then phenylalanine metabolism (9.2507) and finally (glycolysis or gluconeogenesis (8.778), as the most significant pathways in this NPC1 disease dataset. The metabolite pathway decreases from red to yellow then grey/white on Figure 12.

- **Aminoacyl-tRNA Biosynthesis Pathway Analysis**

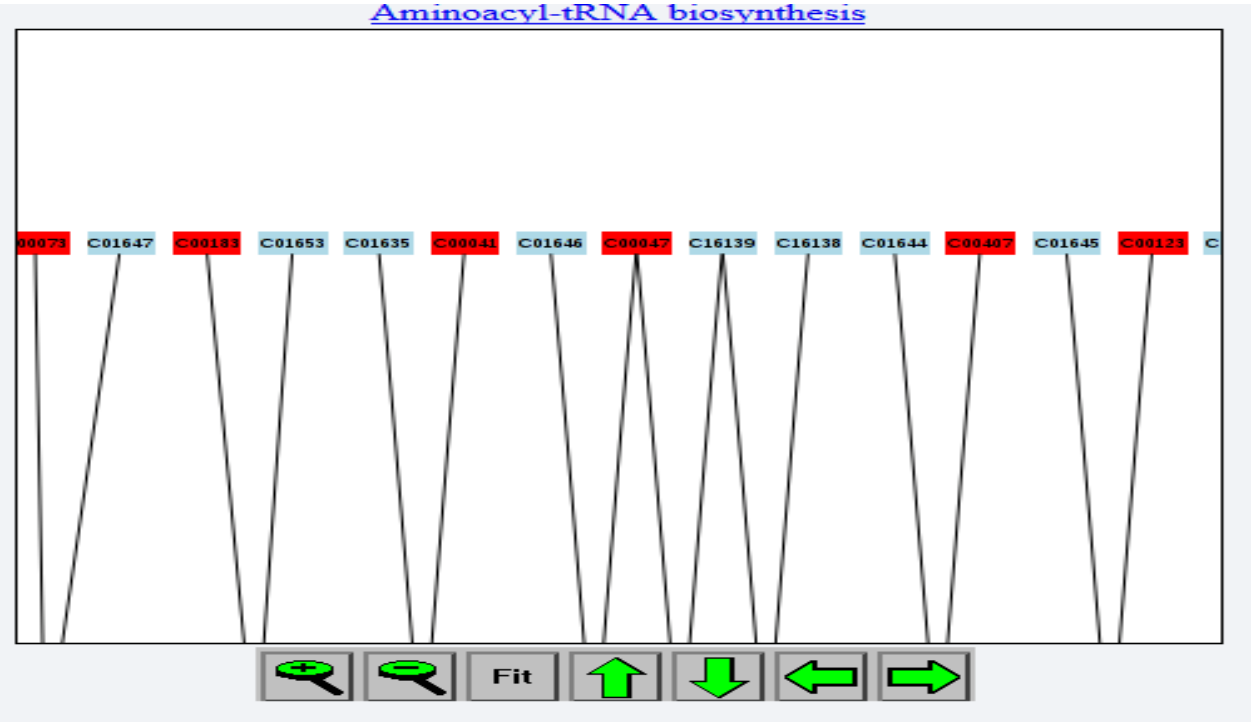

Figure 13. Aminoacyl-tRNA biosynthesis involving in red and from left to right the followed metabolites L-histidine, L- phenylalanine, L- glutamine, **L- methionine, L- valine, L- alanine, L- lysine, L- isoleucine, leucine and L-tyrosine** present in the current blood plasma dataset. The low Holm adjusted p value 2.4815E-5, and the small value of FDR 2.4815E-5 show that the current pathway is strongly expressed in the NPC1 disease aetiology. The ones in bold are the ones shown in red on this Figure and are present in the current dataset.

Several metabolites were observed in the present ^{1}H NMR spectra acquired, and are followed with respect to aminoacyl-tRNA biosynthesis. The L-glutamine can generate ornithine, which in turn produces the pyruvate, with all the three biomolecules present in this pathway. Ornithine can be transformed into the gamma-aminobutyric acid, which is present in the current pathway. This transformation is performed through different intermediate such as the 4-acetamidobutanoate. However, L-aspartic acid can generate creatine. In this pathway analysis, pyruvate, glutamine and glutamate exert the greatest pathway impact, followed by alanine and aspartate. Different pathways are connected to the current pathway, including the

phenylalanine metabolism which is next investigated. The information obtained from Table 27 are further corroborated by the most significant pathways detected and their analysis that follow below (Ruiz-Rodado et al., 2014).

- **Phenylalanine Pathway Analysis**

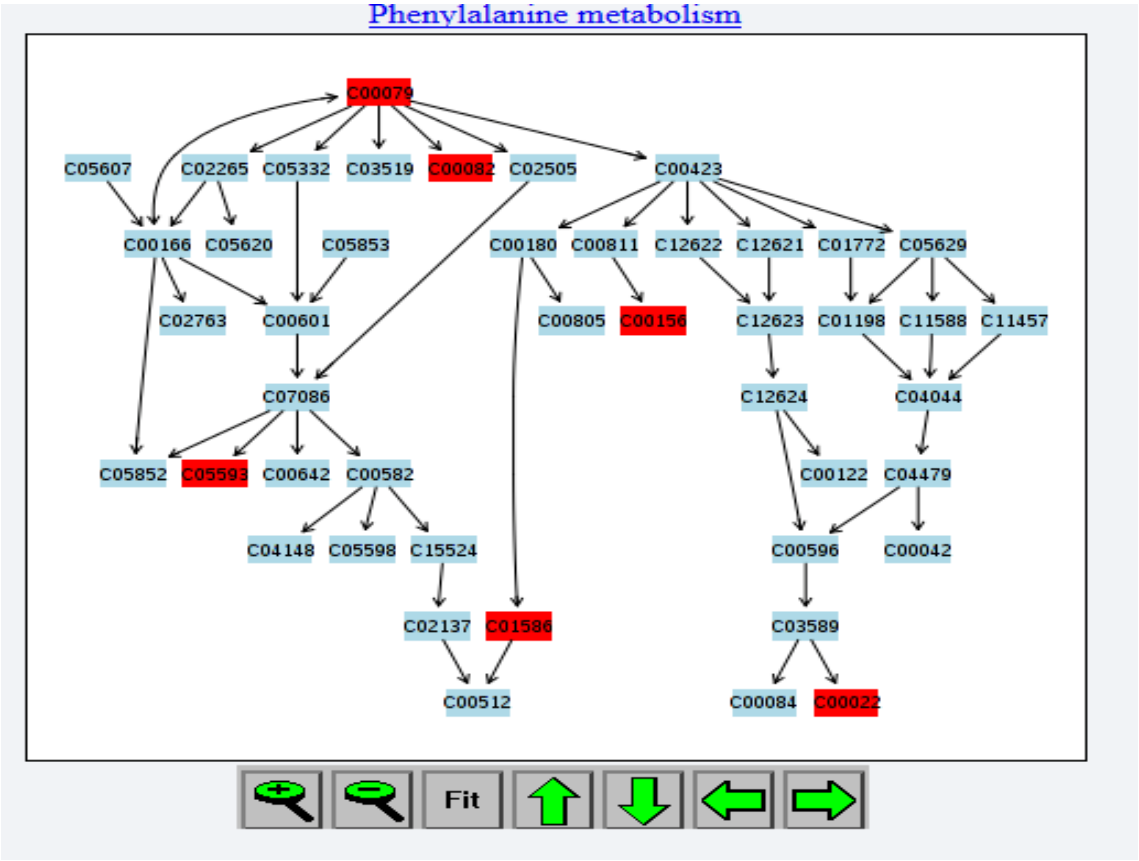

Figure 24. Phenylalanine metabolite pathway analysis with a Holm adjusted p value of 0.0075874 a pathway impact of 0.15 and an FDR of 0.0038. Therefore, the current pathway is strongly differentially expressed in the NPC1 disease dataset, with the presence of the following metabolites, including p-hydroxyphenylacetate, hippurate, pyruvate, fumarate and the succinate.

Phenylalanine metabolism pathway starts with phenylalanine (C00079) itself, and tyrosine (C00082) can be directly generated by oxidation, other metabolites not present in the present dataset allow the generation of 3-hydroxyphenylacetate (C05593), and then the end product on this sub-pathway, which is the benzoyl-CoA. However, phenylalanine can generate trans-cinnamate (C00423), which, in turn can produce hippurate (C01586) through several successive oxidations and hydrolyses. Similarly, trans-cinnamate can produce benzoate

(C00180), which generates the hippurate with one of the highest importance rates 0.0315 for this pathway. Finally, the hippurate can produce the benzoyl-CoA, the terminal product of this pathway analysis.

- **Pathway Analysis of the Glycolysis or Gluconeogenesis**

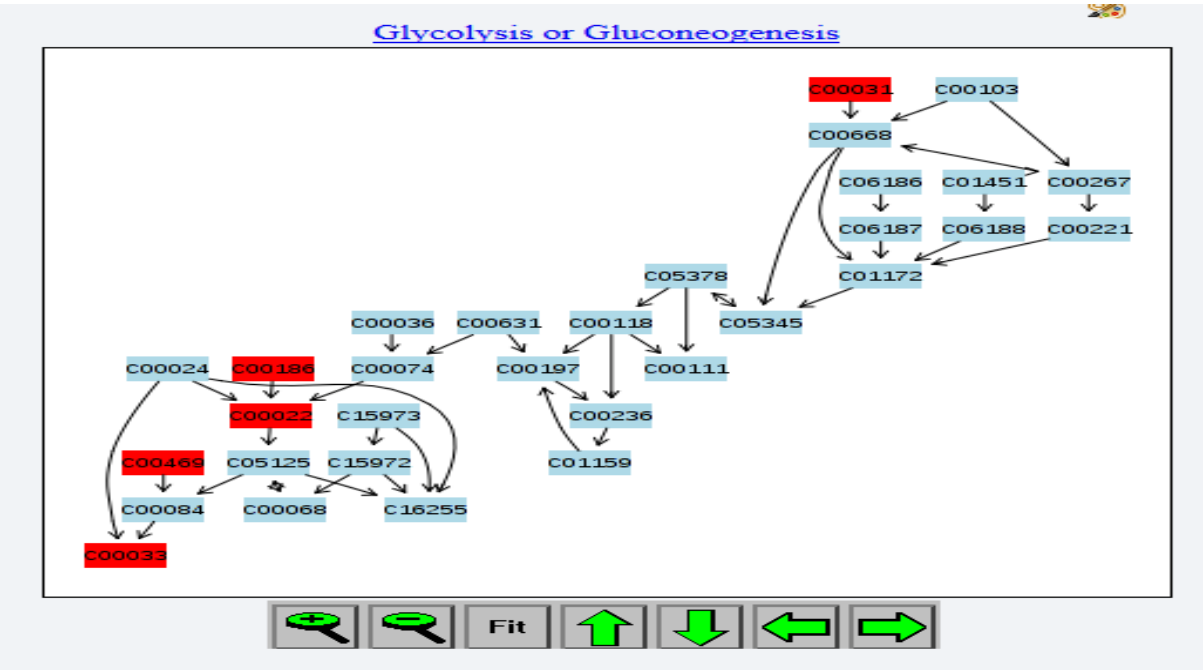

Figure 35. Glycolysis or gluconeogenesis showing in red from the top right corner the sequences of transformation from glucose (C00031), to lactate (C00186), then pyruvate (C00022), followed by ethanol (C00469), and finally on the top bottom left corner to acetate (C00033). Intermediates including alpha-glucose, acetyl-CoA (C00024), acetaldehyde (C00084), (C00074), (C00084), (00068), etc.

However, the matched pathway impact shows that pyruvate, glutamine, and glutamate metabolism exhibit the highest impact on the present glycolysis pathway while the aminoacyl-tRNA biosynthesis shows the highest -log(p) with the smallest Holm adjusted p value and the smallest false detection rate (FDR) in the NPC1 disease dataset.

The glycolysis pathway produces several intermediate metabolites, most of which are not present in this dataset. For example, the intermediate acetyl-CoA (C00024) can generate acetate following several routes as shown above on the Figure 15. The end product of glycolysis is pyruvate. Through successive intermediary metabolites, lactate is generated leading to the

pyruvate (importance = 0.0953) production. Pyruvate serves as an important intermediate in the metabolism of proteins and fats (Pubchem, 2017). It has been detected by the intelligent tri-modelling techniques as one of the main biomarkers in the disease diagnosis and development, i.e ranked 3rd main biomarker (see Table 25). Further transformation can occur with ethanol oxidisation into acetaldehyde before further oxidation to acetate.

5.4. Intelligent Tri-Modelling Techniques and the NPC Liver Dysfunction Disease (NPC LDD) Dataset Analysis – Mouse Model-Based

5.4.1. Introduction

Niemann-Pick type C1 disease is related to the defect of lipid storage at the lysosomal level that end up with fat storage and obesity. More damage can be observed in term of patient's health condition, with cell and tissue death such as liver-related diseases (liver damage). Although such damage may be modulated, the disease can be fatal to some patients (Sayre et al., 2010). Vital and more effective treatments can be administered if NPC disease is discovered early, and hence is potentially circumventable. The use of techniques such as the intelligent tri-modelling techniques (ITMTs) applied in this research can provide answers regarding the discovery of potential biomarkers in the NPC-associated liver dysfunction, more severe form of the NPC1 disease (first study conducted in this thesis). The current intelligent tri-modelling techniques will alternatively involve the scalar visualisation algorithm (SVA), the optimum support vector machine (OSVM) and the principal components regression (PCR) for the scrutiny of this NPC liver dysfunction disease (NPC LDD) dataset, based on mouse model.

5.4.2. Description of the Liver Dysfunction Disease Dataset

- **Description**

The liver dysfunction disease dataset encompasses 65 observations and 143 chemical shifts. They were obtained using NMR spectroscopy on liver extracts collected from mutant mice. The NMR spectrometry generates spectra amplitudes against radio frequency converted into chemical buckets expressed in ppm (part per million). Initially 71 observations were obtained and 6 of them were removed for various reasons. This includes the rows with 0 values (2

observations) and rows with duplicated values (4 observations). The remaining 65 were carried forward for analysis. The statistics obtained and briefly shown below.

Amongst these, 38 were liver extracts samples collected from healthy control mice and the remaining 27 were NPC disease mutant mice. The statistics obtained using XLSTAT are briefly shown below, the full set in included in the appendix section.

- **Statistics**

Different statistics were generated form the XLSTAT software and are below presented.

Summary statistics:							
Variable	Observations	with missing	without missing	Minimum	Maximum	Mean	Std. deviation
[0.77 .. 0.80]	65	0	65	0.000	0.027	0.004	0.004
[0.80 .. 0.82]	65	0	65	0.000	0.035	0.005	0.006
[0.82 .. 0.84]	65	0	65	0.001	0.040	0.007	0.007
[0.84 .. 0.86]	65	0	65	0.002	0.043	0.009	0.008
[0.86 .. 0.92]	65	0	65	0.013	0.097	0.038	0.016
[0.92 .. 0.94]	65	0	65	0.006	0.034	0.015	0.005
[0.94 .. 0.99]	65	0	65	0.025	0.178	0.073	0.031
[0.99 .. 1.05]	65	0	65	0.010	0.096	0.035	0.019
[1.05 .. 1.08]	65	0	65	0.003	0.038	0.013	0.006
[1.13 .. 1.15]	65	0	65	0.001	0.049	0.006	0.008

Table 28. Partial summary statistics obtained from XLSTAT running PCA on the liver dysfunction disease dataset (mice-model). The variables are chemical buckets with the means and standard deviation shown on the right hand-side of the table. The small values of the standard deviation show that the data point are not widely spread out (for this partial summary statistics).

This visualisation of the internal data structure of NPC liver dysfunction disease (mice-model) and the series of developmental processes involved in the data analysis are outlined below.

5.4.3. NPC Liver Dysfunction Disease (NPC LDD) Dataset Analysis using Scalar Visualisation Algorithm (SVA)

- **Liver dysfunction disease dataset analysis by the SVA**

The scalar visualisation algorithm (SVA) as part of the ITMTs will be used for the primary assessment of the NPC liver disease dataset. In order to identify the underlying data structure and detect possible biomarkers that could provide an effective response to the liver failure that occurs in the worst-case scenario. SVA creates a function between the scalar dataset and the selected colours, in such a manner that the current dataset can be visualised while highlighting some more important features (Komura, 2016). The result of this scalar data visualisation process is presented below.

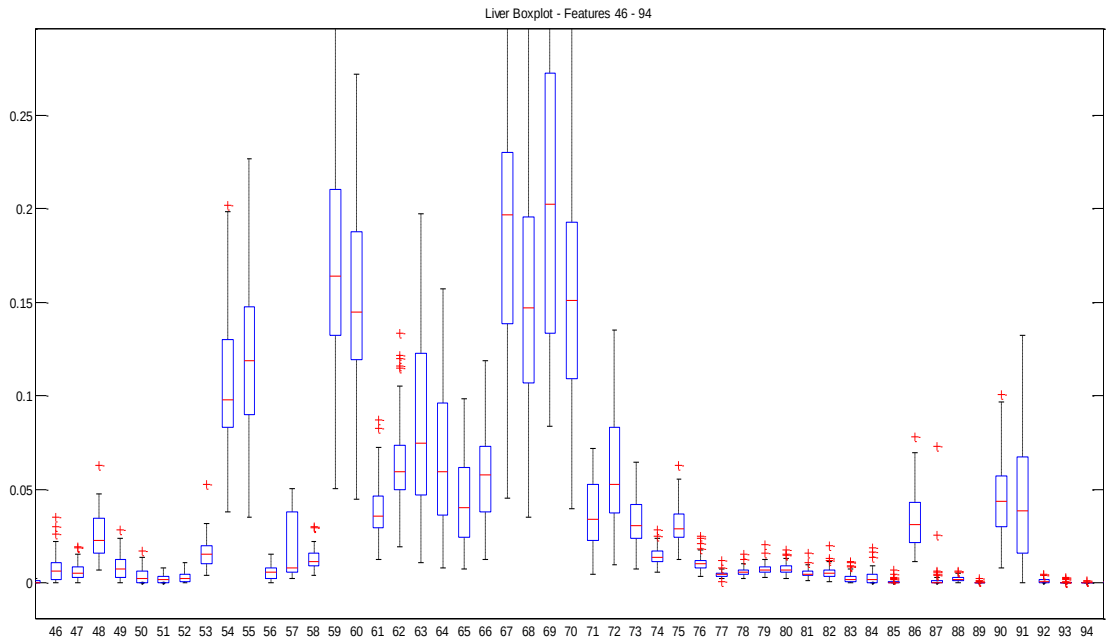

Figure 46. Zooming in the NPC liver disease (mice model) boxplot features showing that the features 69, 67, 59, 70, 68, 60, 55, 54, 63, 62, 64, 66, 72, 65, 90, 91 as potential biomarkers including glutamate, glutamine, taurine, lactate, glycerophosphocholine, myo-Inositol, gamma-phosphorylcholine, leucine, isoleucine, alanine, etc., in the NPC liver dysfunction disease (NPC LDD) diagnosis according to the boxplot technique.

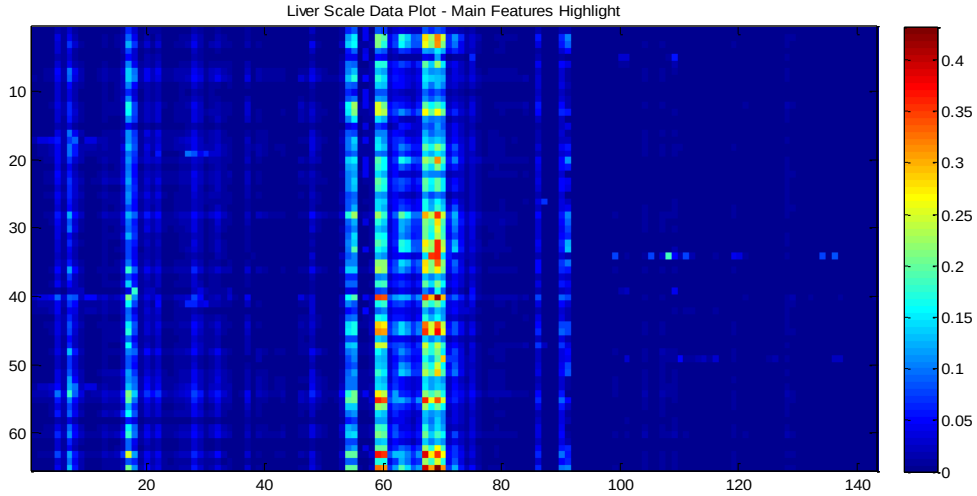

Figure 57. The Scale Data Plot showing significant changes of intensity for the main features 69, 68, 70, 67, 58, 59, 60, 55, 54, 17, 18 – corresponding to the following biomolecules glutamate, glutamine, taurine, lactate, glycerophosphocholine, myo-Inositol, alanine, leucine, and isoleucine considered as potential biomarkers in the liver dysfunction and associated disease diagnosis related to a mouse model, according to this plotting technique.

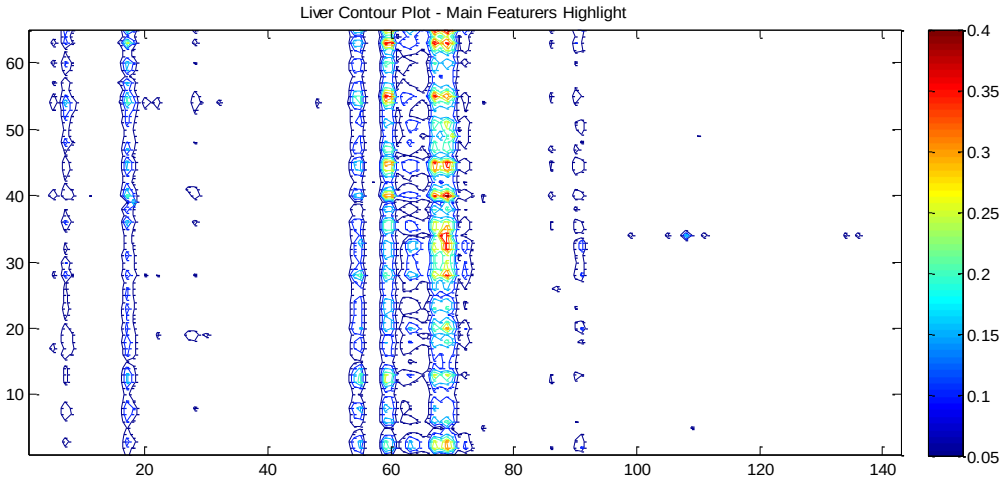

Figure 68. Contour plot showing significant changes of intensity for the most relevant features 69, 68, 70, 67, 55, 54, 17, 18 corresponding respectively to the following biomolecules; glutamate, glutamine, taurine, lactate, glycerophosphocholine, myo-Inositol, alanine, leucine, and isoleucine considered as potential biomarkers in the liver dysfunction disease (mouse model-based) diagnosis according to the contour plot.

- **Results**

The results of the scalar visualisation process have been grouped in the following Table 29 in order to highlight the most important features based on the different plotting techniques.

Plots Names	*Features Number – Important Pattern*	*Corresponding Chemical Features (Shift buckets/ppm)*	*Names of Molecules Detected*
Boxplot	69, 67, 59, 70, 68, 60, 55, 17, 54, 7, 63, 62, 64, 66, 72, 18	[3.81…3.87] **[3.71…3.77]** **[3.39…3.44]** **[3.87…3.93]** **[3.77…3.81]** [3.44…3.50] **[3.26…3.31]** **[1.30…1.36]** **[3.21…3.26]** **[0.94...0.99]** **[3.57…3.62]** **[3.54…3.57]** **[3.62…3.66]** [3.69…3.71] [3.95…3.99] **[1.45...1.51]**	Glutamate, Glutamine, Taurine, Lactate, Glycerophosphocholine, Myo-Inositol, Gamma-phosphorylcholine, Leucine, Isoleucine, Alanine, …
Plot	69, 68, 70, 67, 58, 59, 60, 55, 54, 17	[3.81…3.87] **[3.71…3.77]** **[3.39…3.44]** **[3.87…3.93]** **[3.77…3.81]** [3.44…3.50] **[3.37…3.39]** **[3.26…3.31]** [3.21…3.26] **[1.30…1.36]**	Glutamate, Glutamine, Taurine, Lactate, Glycerophosphocholine, Myo-Inositol, …
Data Scale Plot	69, 68, 70, 67, 58, 59, 60, 55, 54, 17, 18, 7	[3.81….3.87][3.71…3.77] [3.39…3.44] [3.87…3.93] [3.77…3.81] [3.44…3.50] [3.37…3.39] [3.26…3.31] [3.21…3.26] [1.30…1.36] [1.45...1.51] [0.94...0.99]	Glutamate, Glutamine, Taurine, Lactate, Glycerophosphocholine, Myo-Inositol, Alanine, Leucine, Isoleucine, …
Contour Plot	69, 68, 70, 67, 58, 59, 60, 55, 54, 17, 18, 7	[3.81…3.87] [3.71…3.77] [3.39…3.44] [3.87…3.93] [3.77…3.81] [3.44…3.50] [3.37…3.39] [3.26…3.31]	Glutamic acid, Glutamine, Taurine, Lactate, Glycerophosphocholine, Myo-Inositol,

		[3.21…3.26] [1.30...1.36] [1.45...1.51] [0.94...0.99]	Alanine, Leucine, Isoleucine, …

Table 29. SVA Identification patterns of potential biomarkers obtained from several plotting techniques. In green SVA selection of key biomarkers, and in blue fairly important ones. The one in red have not been identified (see Table 52 appendix) although they are important biomarkers in certain plotting techniques such the plot and data scale plot in this mouse-model study.

Features selected from the boxplot representation are those with the amplitude ≥ 0.05. Thus, the different results obtained by the SVA show a great level of consistency with respect to the important features discovered in the different plots.

With regard to the colour mapping algorithm (CMA), the three main plots generated are the plot, the boxplot and the scale data plot, and these show that the chemical shift buckets [3.71…3.77], [3.39…3.44], [3.87…3.93], and [3.77…3.81] ppm corresponding to the to the following features glutamine, glutamate, taurine, and glycerophosphocholine, are very important regarding NPC LDD diagnosis mouse model-based, while those at [3.26…3.31], [3.21..3.26], [3.57…3.62], [3.62…3.66] and [3.69…3.71] ppm corresponding to the biomolecules taurine, myo-inositol, phosphocholine, and gamma-phosphorylcholine are important to a lesser degree.

Although it is difficult to make a clear-cut distinction between the features based on the scale data and contour plots, the liver disease boxplot allows us to do so; a clear distinction is made between the features and their supposed role in understanding the NPC liver disease (mice model) aetiology.

Regarding the contour plot, the result acquired found the same important features, providing a clear understanding of the NPC liver disease development, and also the ones that play a less important role. However, the feature (3.37-3.39) ppm has been detected via the boxplot as a less important feature, whilst in all the other plots it presents as a major feature and it has not been identified (see Table 49). Moreover, a clear determination of the value of certain features can be difficult; however, overall the plotting techniques have suggested potential biomarkers based on the features peaks in the boxplot.

5.4.4. Optimum Support Vector Machine and the NPC Liver Dysfunction Disease (NPC LDD) Data Classification

➢ Data analysis techniques

The analysis of the NPC LDD dataset is quite similar to the one used in the NPC1 disease dataset analysis (paragraph 5.2.3) as they relate to the same disease. In addition, based on the results obtained from the NPC1 patients' blood plasma analysis, the RBF kernel was outperforming the linear kernel and the same results were obtained in similar studies performed elsewhere (Martí and Reinelt, 2011; Tang et al., 2009). And the results of the comparison between the AUC ROC and the Gmean, which show that the former is more appropriate to this research than the latter. Consequently, the application of the OSVM in this second part of this research will be shorten with less tuning process needed, less step in future selection, etc.

Therefore, the result of the present analysis will be focus only on the best OSVM built from precedent data analysis techniques. After confirming through the NPC1 disease dataset, that the optimum SVM (OSVM) model build in the present research is in line with most of the findings related to the OSVM. This includes the RBF kernel outperforming the linear kernel, while the AUC ROC value, when used as performance metrics generates more improved results with the OSVM classifier than the Gmean one (Martí and Reinelt, 2011; Tang et al., 2009).

The optimum support vector machine (OSVM) model built will be used only optimal. Therefore, the analysis in the remainder of this thesis will be using different combinations enabling the normal SVM classifier to be transformed into the OSVM, with their performances being compared against each other. For example, the current dataset was not standardised nor normalised prior used in this model. This was however, included as option in the modelling stage. Therefore, it important and relevant to include the option of standardisation "ON'' (SON) and standardisation "OFF" (SOFF) in the tuning of the OSVM. Overall, the following variables setting were used, including the nearest neighbour for oversampling (NNO), the k folds Used (KFU), the number of repeated stratified cross-validation (RSCV), and the proportion of features used (POFU) for the reliefF algorithm. Furthermore, the nearest neighbour for features selection (NNFS) was applied only in case the variable POFU $\neq 1$.

- **Consideration**

The combination selected for the OSVM model included the choice of the parameter values that optimised the classifier. Consequently, the following parameters setting was adopted. The RBF kernel function was chosen over the linear kernel, whilst, the AUC ROC was selected over the Gmean as a performance metric for improved performance. However, the proportion of features used (POFU) was 1. Therefore, no nearest neighbour features selection (NNFS) was applied. The k folds used was KFU $\geq$ 10, whilst the repeated stratified cross-validation values were set to RSCV $\geq$ 10. The RSCV allows to have a fair representation of each class in each subset of the NPC liver dysfunction dataset used during the cross-validation process (Kim, 2009).

Furthermore, the selection of the parameters C and gamma for tuning the OSVM classifier is performed through the grid search algorithm, which reduces the search space in order to seek local maxima. The different values selected by the algorithm for each local maximum are used for the OSVM dichotomous classification that aims to discriminate between the diseased subjects and the wild treated ones. As a consequence, a set of OSVM classifier performance measures is generated from which an average value is calculated. It should also be noted that the LIBSVM as a stochastic classifier will generate different performance values for different runs of the classifier, and hence the need to determine an average value (McCullagh and Yang, 2006). The OSVM offers potential to discriminate between the NPC liver disease animals and the healthy control ones. Therefore, a real opportunity to predict the NPC liver dysfunction (mouse model-based) and associated disease aetiology at an early stage with higher accuracy, and using the segregation method noted above.

- **OSVM classification results**

In Table 30 the general condition applied involves NNO = 10; KFU = 10; RSCV = 10; POFU = 1; --- RBF and AUC ROC --- (the values in the 1st column relate to that general condition - bold). The modified variables are included.

Modified Variables	SOFF NNO=10	SON NNO=10	SON NNO=15	SON RSCV=15	SON KFU=5
Gamma	**0.239578**	0.001159	0.001281	0.001437	0.001059
CostC	**12570.77**	12.38028	69.37684	12.68029	31.828
AUC ROC	**0.950102**	0.978896	0.978913	**0.979904**	0.978455
Standard Deviation	**0.009304**	0.003783	0.004056	**0.003168**	0.003793
95% Confidence Interval	**±0.018235**	±0.007416	±0.00795	**±0.00621**	±0.007434

Table 30. Modifications of variables and the corresponding increases in the OSVM classifier performances, with RSCV = 15 producing the best AUC ROC value of 0.9799 with minimum standard deviation and confidence interval.

- **Consideration**

The OSVM classifier performances using the standardisation on (SON), the RBF kernel, and the AUC ROC as performance metrics show that increasing the oversampling, the number of repeated stratified cross-validation, the proportion of features used, and the number of K folds employed in the cross-validation process improves the OSVM performance. This result suggests that increasing the different parameters, including NNO, KFU, RSCV, and standardising the dataset increases the classifier power in segregating between the NPC liver disease animals and the wild treated (WT) ones, with higher accuracy gives that the average AUC ROC value is ≥ 0.979, and the average standard deviation is 0.0037 and the confidence intervals fluctuating, although the variation being very small values with an average of ± 0.0073.

A similar result has already been established in previous studies (Martí and Reinelt, 2011; Tang et al., 2009). However, the classification result achieved by the OSVM is far higher than the research results noted above.

5.4.5. Principal Components Regression for the NPC Liver Dysfunction Disease (NPC LDD) Biomarkers Discovery

This principal components regression strategy was implemented in two stages, which included principal components analysis (PCA) to define the principal components (PCs) which are axes orthogonal to each other, and the actual responses/dependents variables Y, which is regressed on these PCs. Each PC is a linear combination of the original variables/features, with each PC_i explaining as much of the remaining variability as possible in the dataset, i.e. they are independent of each other (Anton, 1987; Manfredi, 2013).

- **PCA and the NPC liver dysfunction disease (NPC LDD) dataset**

- **PCA analysis techniques**

In this analysis, none of the principal components/factors were predominant for an explanation of variability. Therefore, a certain number of the PCs is detected to give a sound variability explanation in the NPC LDD. Finally, different levels of variability were considered, including the 90, 80, 70, and the 60% levels in the NPC LDD mouse model dataset. This corresponds to a certain number of PCs necessary to cover the necessary level of variability in the aforementioned dataset. The application of PCA follows the same principle used in section 5.2.4., visited above.

- **Results of the PCA**

The results presented are those really important regarding the remainder of the data analysis, especially for the MLR study that will follow after this section. Based on this reality, the following are presented; this includes the eigenvalues to inform on the level of variability to be considered, and the related number of PCs involved. The factor loadings and the distribution of the features based on the PCs are explored as well. Partial values of PCs, factors loadings, and factor score vectors were presented to allow deeper understanding and the involvement of some of these values, such as the coefficients of correlation and regression, in the selection of the main biomarkers.

- **Principal component analysis:**

- **Partial table of eigenvalues**

PCs	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
Eigenvalue	52.567	19.267	15.312	10.811	7.598	6.660	5.187	3.518	2.853	2.402	2.370	2.024	1.741	1.333
Variability (%)	36.760	13.473	10.708	7.560	5.313	4.658	3.627	2.460	1.995	1.680	1.657	1.415	1.217	0.932
Cumulative %	36.760	50.234	60.942	68.501	73.814	78.472	82.099	84.559	86.555	88.234	89.891	91.307	92.524	93.456

Table 31. Eigenvalues providing variability and the cumulative variability levels. Whereas, only three PCs are needed to explain 60% of the variability, ten are required to explain 90% of the variability. They are used to create biplots in Figure 19 below.

- **Partial Table of factor loadings**

Features	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
[0.77 .. 0.80]	0.747	0.025	-0.112	-0.425	-0.108	0.305	-0.267	0.102	-0.075	0.009	0.038	-0.165	-0.009	0.038
[0.80 .. 0.82]	0.598	0.020	-0.102	-0.378	-0.102	0.420	-0.405	0.155	-0.153	0.065	0.064	-0.195	-0.100	0.064
[0.82 .. 0.84]	0.637	0.020	-0.096	-0.374	-0.098	0.409	-0.396	0.133	-0.134	0.060	0.050	-0.189	-0.097	0.050
[0.84 .. 0.86]	0.624	-0.007	-0.097	-0.331	-0.093	0.450	-0.405	0.120	-0.141	0.077	0.048	-0.191	-0.114	0.048
[0.86 .. 0.92]	0.897	-0.111	0.008	-0.171	-0.038	0.132	-0.209	-0.170	0.058	-0.016	-0.075	-0.036	-0.071	-0.075
[0.92 .. 0.94]	0.932	-0.171	0.022	0.008	0.016	-0.053	-0.118	-0.140	0.049	-0.058	-0.124	0.066	0.050	-0.124
[0.94 .. 0.99]	0.790	-0.278	-0.049	0.331	0.136	-0.059	-0.188	-0.064	0.005	-0.054	-0.185	0.173	0.007	-0.185
[0.99 .. 1.05]	0.782	-0.203	-0.152	0.065	0.054	0.243	-0.412	0.117	-0.084	0.029	-0.099	-0.019	-0.052	-0.099

Table 32. Partial NPC LDD principal components and the contribution of the original features to these PCs.

- **Partial Tables of factor scores**

Observation	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
Obs1	-9.269	-1.709	-3.233	-2.041	-0.554	0.531	1.303	-1.146	-0.823	-0.892	0.141	1.108	-1.075	0.240
Obs2	-3.341	-0.062	7.808	-0.468	-0.791	2.415	2.574	-1.194	0.440	0.004	-0.021	0.693	0.181	-0.876
Obs3	0.327	-0.522	3.431	2.549	0.628	-1.107	0.239	0.260	0.404	0.643	-2.451	2.020	0.065	0.994
Obs4	-9.507	-1.488	1.032	-0.974	-0.932	2.938	1.247	-2.017	0.038	-0.466	0.008	0.020	-0.096	-0.437
Obs5	-8.809	3.733	-7.050	-3.268	-0.676	-1.359	2.622	6.268	6.045	0.433	-2.374	-0.735	-0.404	-3.871
Obs6	-4.401	0.482	0.425	-1.205	-0.204	-2.172	0.527	2.402	-0.268	0.409	-0.439	0.032	-1.085	0.013
Obs7	5.530	-3.627	-7.301	2.109	0.814	-0.764	2.108	-0.345	-0.759	1.196	0.650	0.840	-1.638	0.504
Obs8	9.877	-4.174	-6.913	K0.033	-0.824	0.791	6.813	-0.402	-2.211	-1.590	-0.321	1.065	-0.276	0.136
Obs9	-6.193	-0.446	-1.267	-1.463	-0.488	-1.094	-0.676	-0.379	-0.287	-0.762	-0.029	-0.061	0.067	0.823
Obs10	-8.254	2.074	-3.429	-3.396	-0.239	-0.524	-0.304	-2.012	0.177	-0.723	-0.890	0.008	0.259	-0.049

Table 33. Partial NPC Liver dysfunction disease (NPC LDD)’s factor scores, giving the contribution of each observation to each of the factors/PCs.

- **Biplots**

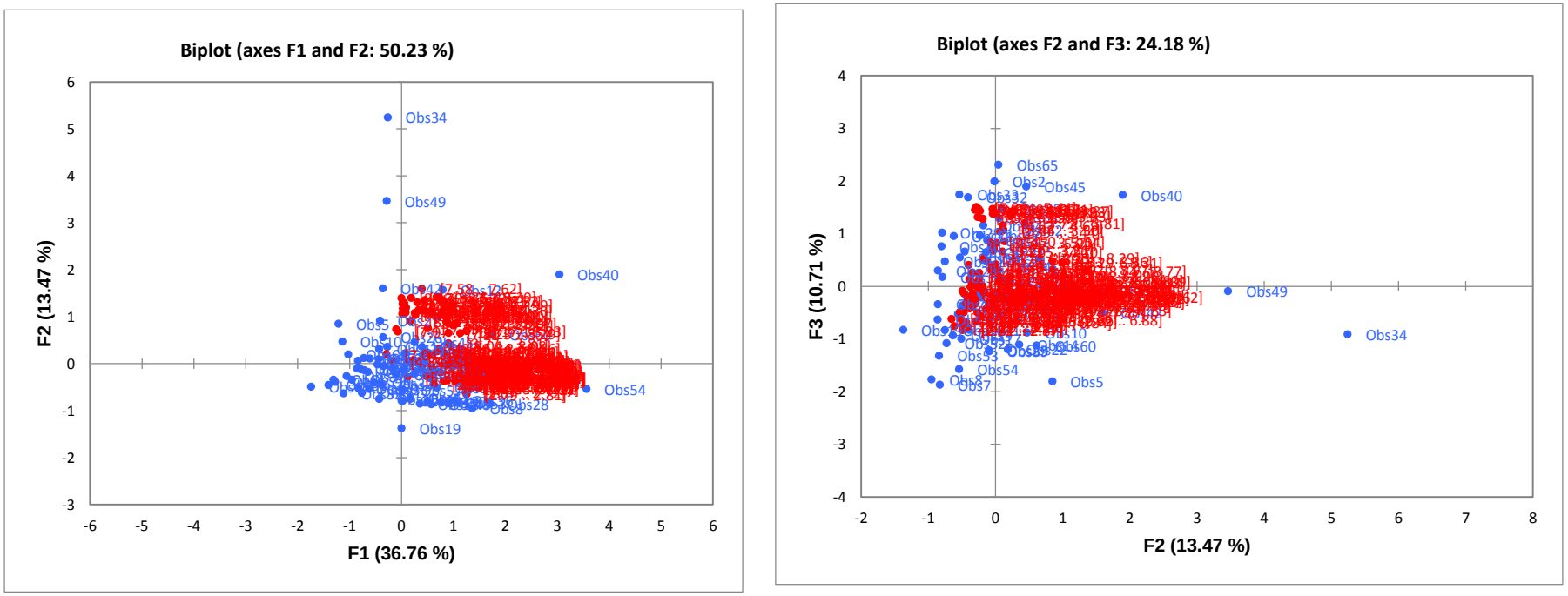

Figure 79. The biplots for factors (F1, F2) and (F2, F3), showing a distinct segregation between the NPC liver disease features, with the red colour representing the diseased group and the blue the wild treated (WT) one, with observation 34 and 49 appearing as outliers.

- **Principal components regression and the NPC liver dysfunction disease (NPC LDD) dataset**

- **Analysis techniques**

  As in section 5.2.4., a correlation exist between the different features in the NPC liver dysfunction disease (NPC LDD) datasets; consequently the PCA and MLR were employed in the following data analysis.

The PCA was used to select the PCs explaining the maximum variability in the original NPC LDD datasets. However, MLR was employed to regress the dependent variable Y on the PCs. Hence, correlating the NPC LDD features to the disease diagnosis and progression. Therefore, a combination of these two techniques allows the implementation of principal components regression (PCR).

Two sets of coefficients including the coefficient of regression (r) and the coefficient of correlation (α) were generated following the implementation of the PCR. They are used to rank the key features and detect biomarkers using the TRTs, similar to section 5.2.4. This includes, the heuristic model, the sum product of the coefficients, and the exponential of the sum of the product of the coefficients' techniques, which subsequently are employed in the biomarkers discovery for the NPC liver dysfunction disease diagnosis.

- **PCR results for the NPC liver dysfunction disease (NPC LDD) dataset**

Based on the percentage of variability, different PCs regression value are obtained. The sum of the product of the regression values (r) by the correlation coefficient values (α) allows to get a final classification for the original features with regards to their importance related to the detection of the potential biomarkers in the NPC LDD using a mouse model. The different statistics related to the model and the features are provided as well.

- **Biomarker discovery in NPC liver dysfunction disease (NPC LDD) with 90% variability**

The results of the ranking obtained, the determination of the potential biomarkers in the NPC1 disease diagnosis for a maximum 90% variability based on the sum of the product of the coefficients, the heuristic parameter, and the exponential sum of the product of the coefficient methods are given in the following Tables, with some statistics related to the method and the PC.

Liver Dysfunction Disease 90% Variability				
Model Statistics	**Coefficient**	**SE**	**t-statistics**	**p-value**
Deviance of the Fit:	7.5737	null	Null	Null
Degrees of Freedom:	52	null	Null	Null
Estimated dispersion parameter:	0.423	null	Null	Null
Feature Number				
0	**-5.6936**		Statistically insignificant	>0.05
1	**-2.0553**		"	>0.05
2	**1.5886**		"	>0.05
3	**-6.7296**		"	>0.05

4	**11.1199**		"	>0.05
5	**11.9522**		"	**>0.05**
6	**16.3454**		"	>0.05
7	3.392		"	>0.05
8	4.4974		"	>0.05
9	**-9.2118**		"	>0.05
10	**-3.8479**		"	>0.05
11	**-14.3557**		"	>0.05
12	**-5.2846**		"	>0.05
***Feature number 0 is the intercept.**	Null	null	Null	Null
Null	Null	null	Null	Null

Table 34. Principal components and the related coefficients of regression and associated statistics such as the standard error, and the p values for the 90% variability. The ranking obtained based on the different statistics provided above and the tri-ranking techniques (TRTs) follows below. The percentage of selections are noted correct ranking = CR and incorrect ranking = IR in comments column.

Ranking Method	**Potential Biomarkers (90% Var.)**	**Comment**
Heuristic Method	[7.47…7.52] [8.12….8.15] [7.65…7.70] [7.72…7.78] [7.62…7.65] [8.08…8.12] [8.69…8.74] [4.47…4.53] [2.59…2.62] [7.07…7.12] [2.62...2.67] [3.93…3.95] [3.95...3.99] [7.30...7.35] [7.70...7.72]	Compared to SPC CR =18/24 = 75% IR =6/24 =25% In red are the biomarkers not detected by the model

	[4.16…4.21] [1.13…1.15] [1.15…1.17] [7.41…7.47] [0.99…1.05] [4.30…4.35] [4.24…4.27] [4.61…4.63] [7.17…7.23]	
Sum of the Product of the Coefficients (SPC) Method	[7.47…7.52] [8.12…8.15] [7.65…7.70] [7.72…7.78] [8.08…8.12] [7.62…7.65] [8.69…8.74] [7.07…7.12] [2.62…2.67] [7.70…7.72] [4.24…4.27] [0.99…1.00] [2.99…3.05] [2.67…2.72] [3.05…3.11] [4.16…4.21] [7.30…7.35] [7.17…7.23] [0.94…0.99] [4.30…4.35] [1.70…1.74] [4.47…4.53] [8.15…8.17] [2.79…2.81]	**Reference Patterns** In green the 24 most effective biomarkers detected by the SPC ranking techniques
Exponential Of the Sum of the Product of the Coefficients (ESPC) Method	[7.47…7.52] [8.12…8.15] [7.65…7.70] [7.72…7.78] [8.08…8.12] [7.62…7.65] [8.69…8.74] [7.07…7.12] [2.62…2.67] [7.70…7.72] [4.24…4.27] [0.99…1.05] [2.99…3.05] [2.67…2.72] [3.05…3.11] [4.16…4.21] [7.30…7.35] [7.17…7.23] [0.94…0.99] [4.30…4.35] [1.70…1.74] [4.47…4.53] [8.15…8.17] [2.79…2.81]	Correct Classification Compared to SPC CR = 24/24 = 100% IR = 0/24 = 0% Same ranking ability as the SPC method

Table 35. Potential biomarkers based on the three different techniques involved above. The tri-ranking techniques (TRTs) generally gives rise to the same selection for the main biomarkers in the NPC liver disease dataset. The 10 most important biomarkers out of 24 are identical (in blue) for the tri-ranking techniques, with different performance especially for heuristic model.

Based on the percentages of variability accounted for, the different PC regression value are obtained. The sum of the product of the regression values (r) via the correlation coefficient values (α) allows us to attain a final classification for the original features with regard to their importance in detection of the potential biomarkers. Nonetheless, the results of the ranking obtained shows that determinations of potential biomarkers in the NPC LDD diagnosis for up to 90% variability, based on the heuristic model, the sum of the product of the coefficients (SPC), and the exponential methods are consistent. However, identical major biomarkers are featured. The heuristic and the SPC provide the same rankings. Potential biomarkers suggested by the SVA are detected by the ranking models.

- **Biomarker discovery in NPC liver dysfunction condition with 80% variability**

The results of the ranking obtained and determination of the potential biomarkers for the 80% variability model, based on the double coefficients, the heuristic, and the exponential methods are provided in the following tables, together with statistics related to the model and its features.

Liver Dysfunction Disease 80% Variability				
Model Statistics	***Coefficient***	***SE***	***t-statistics***	***p-value***
Deviance of the Fit:	27.5949	Null	Null	null
Degrees of Freedom:	57	Null	Null	null
Estimated dispersion parameter:	0.758	Null	Null	null
Feature Number				
0	-1.0643			>0.05
1	-1.0626			>0.05
2	2.8073			>0.05
3	**-2.0675**	**0.9604**	**-2.1527**	**0.0313**
4	**7.6167**	**2.7247**	**2.7955**	**0.0052**
5	-3.3453			>0.05
6	**2.6602**	**1.0261**	**2.5925**	**0.0095**
7	-0.4302			>0.05
*** Feature number 0 is the intercept.**	null	Null	null	null
Null	Null	Null	null	null

Table 36. Different PCs and related coefficients regression, and correlation and associated statistics in the model 80% variability. The small deviance of the fit and estimated dispersion parameter informs of the probable importance of this 80% variability level. The ranking obtained is based on the statistic values provided above and the TRT.

Ranking Method	**Potential Biomarkers (80% Var.)**	**Comment**
Heuristic Method	[6.88…6.93] [7.35…7.41] [7.52…7.58] [7.91…7.94] [7.23…7.28] [8.39…8.44] [7.85…7.91] [8.47…8.53] [7.30…7.35] [2.59…2.62] [7.41…7.47] [7.28…7.30]	Compared to SPC CR =20/24 = 83.33% IR =4/24 =16.67%

	[6.86…6.88] [7.58…7.62] [8.36…8.39] [7.70…7.72] [3.93…3.95] [3.77…3.81] [2.62…2.67] [7.72…7.78] [7.62…7.65] [2.67…2.72] [2.79…2.81] [3.95…3.99]	In red are the biomarkers not detected by the model
Sum of the Product of the Coefficients (SPC) Method	[6.88…6.93] [7.35…7.41] [7.52…7.58] [7.91…7.94] [7.23…7.28] [8.39…8.44] [8.47…8.53] [7.85…7.91] [7.30…7.35] [2.59…2.62] [7.41…7.47] [7.58…7.62] [7.28…7.30] [8.36…8.39] [6.86…6.88] [8.44…8.47] [7.70…7.72] [7.47…7.52] [3.77…3.81] [3.93…3.95] [2.99…3.05] [2.62…2.67] [7.72…7.78] [7.94…7.99]	**Reference Patterns** In green are the 10 most effective biomarkers detected based on the ranking technique
Exponential Of the Sum of the Product of the Coefficients (ESPC) Method	[6.88…6.93] [7.35…7.41] [7.52…7.58] [7.91…7.94] [7.23…7.28] [8.39…8.44] [8.47…8.53] [7.85…7.91] [7.30….7.35] [2.59…2.62] [7.41…7.47] [7.58…7.62] [7.28…7.30] [8.36…8.39] [6.86…6.88] [8.44…8.47] [7.70…7.72] [7.47…7.52] [3.77…3.81] [3.93…3.95] [2.99…3.05] [2.62…2.67] [7.72…7.78] [7.94…7.99]	Correct Classification Compared to SPC CR = 24/24 = 100% IR = 0/24 = 0% In green are the 10 most effective biomarkers detected based on the ranking technique

Table 37. 24 most important buckets corresponding to the 24 best markers (in green) detected by the tri-ranking techniques and potential biomarkers for the 80% total variability dataset. Potential biomarkers based on three techniques applied. In general, the tri-ranking techniques selected the same features as major biomarkers in the NPC liver disease dataset. The 10 most important biomarkers out of 24 are identical for the tri-ranking techniques, with different performance especially for heuristic model (in blue in SPC and ESPC sections).

NB: In orange biomolecules non-identified.

Results for a maximal 80% variability are very similar to those obtained for 90% variability. Indeed, most of the potential biomarkers discovered in both cases are the same, although there are differences in the rankings.

- **Biomarker discovery in NPC liver disease with a maximum70% variability**

The results of the rankings obtained, and the identification of potential biomarkers in the NPC1 disease diagnosis for a total 70% variability, which is based on the coefficients of regression and the coefficient of correlation, the heuristic, the SPC and the ESPC methods, are given in the following Tables 38, and 39 together with statistics related to the model and the features.

NPC Liver Dysfunction Disease - 70% Variability				
Model Statistics	***Coefficient***	***SE***	***t-statistics***	***p-value***
Deviance of the Fit:	39.1255	Null	null	Null
Degrees of Freedom:	59	Null	null	Null
Estimated dispersion parameter:	1.205	Null	null	Null
Feature Number				
0	-1.4455	0.5227	-2.7657	0.0057
1	-0.7047			>0.05
2	-1.5026			>0.05
3	-0.8017			>0.05
4	**3.8655**	**1.0493**	**3.684**	**0.0002**
5	0.403			>0.05
*** Feature number 0 is the intercept.**	null	Null	null	Null
Null	null	Null	null	Null

Table 38. PCs and their related coefficients of regression and associated statistics for the 70% variability model. The fairly higher deviance of the fit and estimated dispersion parameters express the probable low importance of the 70% level of ranking acquired. Feature number 4 in green is the only one to be statistically significant with a p value of value 0.0002 < 0.05.

Ranking Method	**Potential Biomarkers (70% Var.)**	**Comment**
Heuristic Method	[7.30…7.35] [2.62…2.67] [2.79…2.81] [2.67…2.72] [6.88…6.93] [7.85…7.91] [2.32…2.37] [7.35…7.41] [7.52…7.58] [2.81…2.87] [7.91…7.94] [8.47….53] [2.59…2.62] [3.05…3.11] [7.23…7.28] [8.39…8.44] [0.94…0.99] [2.99…3.05]	Compared to SPC CR = 18/24 = 75% IR = 6/24 = 25 % In red are the biomarkers not detected by the model

	[3.95…3.99] [3.93…3.95] [2.37…2.42] [1.70…1.74] [3.31…3.35] [7.17…7.23]	
Sum of the Product of the Coefficients (SPC) Method	[2.62…2.67] [2.79…2.81] [2.67…2.72] [7.30…7.35] [2.81…2.87] [2.32…2.37] [2.59…2.62] [3.05…3.11] [0.94…0.99] [2.99…3.05] [3.31…3.35] [1.70…1.74] [2.37…2.42] [3.11…3.15] [3.15…3.17] [7.17…7.23] [6.88…6.93] [2.26…2.32] [3.95…3.99] [7.85…7.91] [7.41…7.47] [2.72…2.77] [7.35…7.41] [4.16…4.21]	Reference Patterns In green are the 24 best biomarkers detected based on ranking criteria
Exponential Of the Sum of the Product of the Coefficients (ESPC) Method	[2.62…2.67] [2.79…2.81] [2.67…2.72] [7.30…7.35] [2.81…2.87] [2.32…2.37] [2.59…2.62] [3.05…3.11] [0.94...0.99] [2.99…3.05] [3.31…3.35] [1.70…1.74] [2.37…2.42] [3.11…3.15] [3.15…3.17] [7.17…7.23] [6.88…6.93] [2.26…2.32] [3.95…3.99] [7.85…7.91] [7.41…7.47] [2.72…2.77] [7.35…7.41] [4.16…4.21]	Correct Classification Compared to SPC CR = 24/24 = 100% IR = 0/24 = 0%

Table 59. Potential biomarkers based on the three main features ranking techniques formally identified. The ranking techniques select the same features as major biomarkers in the NPC liver disease diagnosis.

Results for 70% maximum variability model are very similar to the one obtained with 90% variability. Indeed, most of the potential biomarkers discovered in both cases are the same, differences observed relating only to rankings. In addition, 70% could not detect all the biomarkers selected by the reference ranking technique, and includes variation in ranking with regard to common biomarkers detected, including the 10 most important biomarkers out the 24 main ones (in blue).

- **Biomarker discovery in NPC liver disease with 60% variability**

The results of the ranking obtained, determination of the potential biomarkers in the NPC liver disease dataset for a model with a maximum 60% variability is shown in Table 40 below. This is based on the coefficients of correlation and regression, the heuristic, the SPC and the ESPC methods, its significant biomarkers together with statistics related to the model and its features.

NPC Liver Dysfunction Disease - 60% Variability				
Model Statistics	***Coefficient***	***SE***	***t-statistics***	***p-value***
Deviance of the Fit:	83.2905	Null	Null	Null
Degrees of Freedom:	61	Null	Null	Null
Estimated dispersion parameter:	1.105	Null	Null	Null
Feature Number				
0	-0.5652	0.2653	-2.1305	0.0331
1	0.0952			>0.05
2	-0.4132			>0.05
3	-0.1481			>0.05
***Feature number 0 is the intercept.**	Null	Null	Null	Null
Null	Null	Null	Null	Null

Table 40. PCs and related coefficients of regression, in addition to different statistics for the 60% variability model. The higher deviance of the fit and estimated dispersion parameters express the probable low importance of the 60% level of ranking acquired. Feature 1, 2, 3 in are less statistically significant with a p value > 0.05.

The ranking obtained from application of the TRT are in Table 41 below.

Ranking Method	**Potential Biomarkers (60% Var.)**	**Comment**
Heuristic Method	[2.67…2.72] [2.79…2.81] [2.72…2.77] [2.81…2.87] [3.05…3.11] [2.26…2.32] [3.15…3.17] [3.11…3.15] [2.51…2.53] [2.87…2.90] [3.31…3.35] [2.62…2.67] [2.37…2.42] [1.05…1.08] [3.17…3.19] [1.70…1.74] [1.74…1.79] [1.95…1.97] [0.94…0.99] [1.89…1.95] [1.97…1.99] [2.32…2.37] [2.01…2.04] [0.99…1.05]	Ranking Performance compared to that of the SPC CR = 24/24 = 100% IR = 0/24 =0 %
Sum of the Product of the Coefficients (SPC)	[2.67…2.72] [2.79…2.81] [2.72…2.77] [2.81…2.87] [3.05…3.11] [2.26…2.32] [3.15…3.17] [3.11…3.15] [2.51…2.53]	**Reference Patterns** In green are the 24 most effective biomarkers

Method	[2.87…2.90] [3.31…3.35] [2.62…2.67] [2.37…2.42] [1.05…1.08] [3.17…3.19] [1.70…1.74] [1.74…1.79] [0.94…0.99] [1.95…1.97] [2.32…2.37] [1.89…1.95] [1.97…1.99] [2.01…2.04] [0.99…1.05]	detected based on the ranking criteria
Exponential Of the Sum of the Product of the Coefficients (ESPC) Method	[2.67…2.72] [2.79…2.81] [2.72…2.77] [2.81…2.87] [3.05…3.11] [2.26…2.32] [3.15…3.17] [3.11…3.15] [2.51…2.53] [2.87…2.90] [3.31…3.35] [2.62…2.67] [2.37…2.42] [1.05…1.08] [3.17…3.19] [1.70…1.74] [1.74…1.79] [0.94…0.99] [1.95…1.97] [2.32…2.37] [1.89…1.95] [1.97…1.99] [2.01…2.04] [0.99…1.05]	**Ranking Performance** compared to that of the SPC CR = 24/24 = 100% IR = 0/24 = 0%

Table 41. Potential biomarkers based on the three main features ranking techniques formally identified for a maximum 60% variability level. These ranking techniques selected the same 24 main potential biomarkers. The ranking performance for heuristic and ESPC methods have been compared to that of the SPC used as ranking reference model. The 10 most important biomarkers of 24 are identical for the three techniques.

Results for this 60% variability model are that TRTs performed an extreme high level of ranking, with the same selection of 24 most effective biomarkers. This can be explained by the low level of variability, i.e. 60% variability only explained. These features ranked are related to the four different levels of variability that are compared below.

- **Comparing the four levels of variability**

Table below is obtained from the precedent features ranking, and using the SPC method as reference for all the variability levels.

90%	80%
[7.47…7.52] [8.12…8.15] [7.65…7.70]	[6.88…6.93] [7.35…7.41] [7.52…7.58]
[7.72…7.78] [8.08…8.12] [7.62…7.65]	[7.91…7.94] [7.23…7.28] [8.39…8.44]
[8.69…8.74] [7.07…7.12] [2.62…2.67]	[8.47…8.53] [7.85…7.91] [7.30…7.35]
[7.70…7.72] [4.24…4.27] [0.99…1.0]	[2.59…2.62] [7.41…7.47] [7.58…7.62]
[2.99…3.05] [2.67…2.72] [3.05…3.11]	[7.28…7.30] [8.36…8.39] [6.86…6.88]
[4.16…4.21] [7.30…7.35] [7.17…7.23]	[8.44…8.47] [7.70…7.72] [7.47…7.52]
[0.94…0.99] [4.30…4.35] [1.70…1.74]	[3.77…3.81] [3.93…3.95] [2.99…3.05]
[4.47…4.53] [8.15…8.17] [2.79…2.81]	[2.62…2.67] [7.72…7.78] [7.94…7.99]
70%	**60%**
[2.62…2.67] [2.79…2.81] [2.67…2.72]	[2.67…2.72] [2.79…2.81] [2.72…2.77]
[7.30…7.35] [2.81…2.87] [2.32…2.37]	[2.81…2.87] [3.05…3.11] [2.26…2.32]
[2.59…2.62] [3.05…3.11] [0.94…0.99]	[3.15…3.17] [3.11…3.15] [2.51…2.53]
[2.99…3.05] [3.31…3.35] [1.70…1.74]	[2.87…2.90] [3.31…3.35] [2.62…2.67]
[2.37…2.42] [3.11…3.15] [3.15…3.17]	[2.37…2.42] [1.05…1.08] [3.17…3.19]
[7.17…7.23] [6.88…6.93] [2.26…2.32]	[1.70…1.74] [1.74…1.79] [0.94…0.99]
[3.95…3.99] [7.85…7.91] [7.41…7.47]	[1.95…1.97] [2.32…2.37] [1.89…1.95]
[2.72…2.77] [7.35…7.41] [4.16…4.21]	[1.97…1.99] [2.01…2.04] [0.99…1.05]

Table 42. Major potential biomarkers detected at different levels of variability with the selection 90% used as variability reference. In green are the features detected by all the four levels of variability, in blue are the ones detected by three levels of variability, and in red are those detected by two levels of variability and in black those detected by only one level of variability. It should be noted that only **(2.62-2.67) ppm** corresponding to methionine and hypotaurine was detected by the 4 levels of variability amongst the 24 best biomarkers. 90% variability could detect most of the biomarkers detected by the other three levels of variability. This includes nicotinate, xanthine, lysine, ornithine, Citrate, phenylalanine, leucine/isoleucine, and aspartate which have been classified as very important biomarkers in the NPC1 disease diagnosis (see Tables 42). The two sets of findings are in line with the fact that methionine and hypotaurine are the most important biomarkers detected by the intelligent tri-modelling techniques developed (ITMTs) are they are ranked No1 in terms of potential biomarkers detected in the NPC liver dysfunction dataset (identified as probable new biomarkers). The other potential biomarkers detected including nicotinate, xanthine, lysine, ornithine, Citrate, phenylalanine, leucine/isoleucine, and aspartate are ranked amongst the 15 most effective biomarkers detected by the ITMTs (Table 43 below).

- **Comparison of models and the ranking methods based on the level of variability**

Level of Variability (%)	ESPC Method (%)	Heuristic Method with CR (%)	Model Deviance of Fit	Parameters Estimated Dispersion	Degree of Freedom (DoF)
90	100	75	7.5737	0.423	52
80	100	83.33	27.5949	0.758	57
70	100	75	39.1255	1.205	59
60	100	100	83.2905	1.105	61

Table 63. Percentage of the correct ranking (CR) performed by the Heuristic method and associated statistics related to each level of variability.

- **Consideration**

1. The Heuristic ranking performance involving the level of correct (CR) is used to compare models as well as the level of variability.
2. The SPC and ESPC ranking methods achieved the same level of rankings were considered as the best ranking techniques based on these criteria.
3. The model deviance and the parameters' estimated dispersion established that 90% and then 80% respectively serve as the best and second-best variability percentages in the NPC LDD diagnosis dataset. This is ascribable to the minimum level of deviance ($7.57 < 27.59$) and the estimated dispersion ($0.423 < 0.758$) for 90% and 80 % variabilities respectively. However, based on the heuristic method which is the ranking technique with the lowest ranking ability, the level of correct ranking performance (CR) for the 80% variability model appears to offer improvements over the 90% level one. In addition, ranking being the main objective in this section, therefore, 80% is considered to be the best level of variability followed by 90% variability although the latter has better statistical parameters. This result is consistent with the one obtained for the NPC1 disease dataset human model-based.

The Table below gives the full range of the 20 main potential biomarkers discovered by the PCR model developed in this study; they are compared to existing biomarkers.

- **Related molecular assignment - 20 most effective biomarkers:**

Ranking	Chemical Shift	Probable Biomarkers in NPC Liver Dysfunction Disease Mouse-Model	Comments on Biomarkers Discovered
1	[2.62…2.67]	Methionine / Hypotaurine	Probable Biomarkers
2	[2.99…3.05]	Lysine / Ornithine	Existing Biomarkers
3	[7.30…7.35]	Phenylalanine	Existing Biomarker
4	[7.47…7.52]	Phenylalanine	Existing Biomarker
5	***[7.72…7.78]***	Hippurate-C4-CH	Probable New Biomarker
6	***[7.70…7.72]***	Nicotinate / Xanthine	Probable New Biomarkers
7	[2.67…2.72]	Citrate-CH_2A / CH_2B	Probable Biomarker
8	[4.16…4.21]	Phosphorocholine	Probable Biomarker
9	[7.17…7.23]	Tyrosine	Existing Biomarker
10	[0.94…0.99]	Leucine-CH_3 / Isoleucine	Existing Biomarker
11	[1.70…1.74]	Lysine-C5-CH_2	Existing Biomarker
12	[2.79…2.81]	Aspartate	Probable Biomarker
13	[2.81…2.87]	Aspartate	Probable Biomarker
14	[2.32…2.37]	Pyruvate-CH_3; Glutamate-C3-CH_2	Existing Biomarkers
15	[3.05…3.11]	Ornithine-CS	Probable Biomarker
16	[3.31…3.35]	Cystine-C3/C6-CH_2	Probable Biomarker
17	[1.70…1.74]	Lysine-C5-CH_2	Probable Biomarker
18	[2.37…2.42]	Glutamate-C3-CH_2; Succinate-CH_2's	Existing/Probable Biomarkers
19	[3.11…3.15]	Glucuronate- C2-CH	Probable Biomarker
20	[3.15…3.17]	9-Methyluric acid	Probable Biomarker

Table 44. Full list of 20 potential biomarkers with probable new biomarkers discovered by the PCR model in the NPC liver dysfunction disease dataset mouse model. Hippurate (5^{th}) and the mixture of nicotinate and xanthine (6^{th}) as probable new biomarkers in the NPC LDD diagnosis.

➢ **Consideration**

For the NPC liver dysfunction dataset mouse-model, the tri-ranking techniques employed have discovered the same main biomarkers overall, However, the top 10 differ from one level of variability to the next. However, for the four different variabilities, the rankings obtained are the same for SPC and the ESPC. The list of the 7 main biomarkers formally detected by all four levels of variability is methionine, hypotaurine, lysine, ornithine, phenylalanine, hippurate, nicotinate, xanthine, and Citrate.

Following the detection of the 7 main biomarkers, their relation to our understanding of NPC1 liver disease allows to carry out a pathway analysis that is next to come.

5.4.6. Biomarkers Pathway Analysis of the NPC Liver Dysfunction Disease

➢ **Result of pathway analysis**

- **Table of major metabolomics pathway in the NPC LDD**

Pathway Name	Match Status	P	-log(p)	Holm p	FDR	Impact	Details
Alanine, aspartate and glutamate metabolism	9/24	5.4248E-9	19.032	**4.4484E-7**	**4.4484E-7**	0.7194	KEGG
Aminoacyl-tRNA biosynthesis	11/69	1.6168E-6	13.335	**1.3096E-4**	**6.6287E-5**	0.0	KEGG
Arginine and proline metabolism	8/44	1.9994E-5	10.82	**0.0015996**	**5.4652E-4**	0.24479	KEGG
D-Glutamine and D-glutamate metabolism	3/5	2.1629E-4	8.4389	**0.017087**	**0.004434**	1.0	KEGG
Nitrogen metabolism	3/9	0.0016751	6.3919	0.13066	0.027472	0.0	KEGG
Citrate cycle (TCA cycle)	4/20	0.0020922	6.1695	0.1611	0.028594	0.17794	KEGG
Butanoate metabolism	4/22	0.0030286	5.7996	0.23018	0.032388	0.02899	KEGG

Valine, leucine and isoleucine biosynthesis	3/11	0.0031599	5.7572	0.23699	0.032388	0.66666	KEGG
Phenylalanine, tyrosine and tryptophan biosynthesis	2/4	0.0047257	5.3547	0.3497	0.043057	1.0	KEGG
Glycine, serine and threonine metabolism	4/31	0.010784	4.5297	0.78726	0.088432	0.26884	KEGG
Ascorbate and aldarate metabolism	2/9	0.025861	3.655	1.0	0.19278	0.4	KEGG
Glutathione metabolism	3/26	0.036837	3.3012	1.0	0.23794	0.06107	KEGG
Phenylalanine metabolism	2/11	0.038091	3.2678	1.0	0.23794	0.40741	KEGG
Cysteine and methionine metabolism	3/27	0.040624	3.2034	1.0	0.23794	0.10977	KEGG
Glycerophospholipid metabolism	3/30	0.053122	2.9352	1.0	0.2904	0.09074	KEGG
Histidine metabolism	2/15	0.06762	2.6938	1.0	0.34655	0.0	KEGG
Ubiquinone and other terpenoid-quinone biosynthesis	1/3	0.084374	2.4725	1.0	0.40698	0.0	KEGG
Lysine biosynthesis	1/4	0.11092	2.1989	1.0	0.50532	0.0	KEGG
Purine metabolism	4/68	0.12964	2.043	1.0	0.52605	0.15984	KEGG
Tyrosine metabolism	3/44	0.13153	2.0285	1.0	0.52605	0.14045	KEGG

Table 45. Metabolomics pathway analysis results derived from the identification of biomarker metabolites in NPC mouse model liver dysfunction process datasets facilitate our understanding of disease aetiology. Based on the Holm adjusted p values (p value < 0.05) and the false detection rate (FDR< 0.05), it is clear that the first metabolomics pathways with the smallest Holm adjusted p-values are the ones that are strongly differentially expressed. The 4

first lines corresponding to major pathways appear to be significantly important in this process are further discussed below.

- **Pathway Overview**

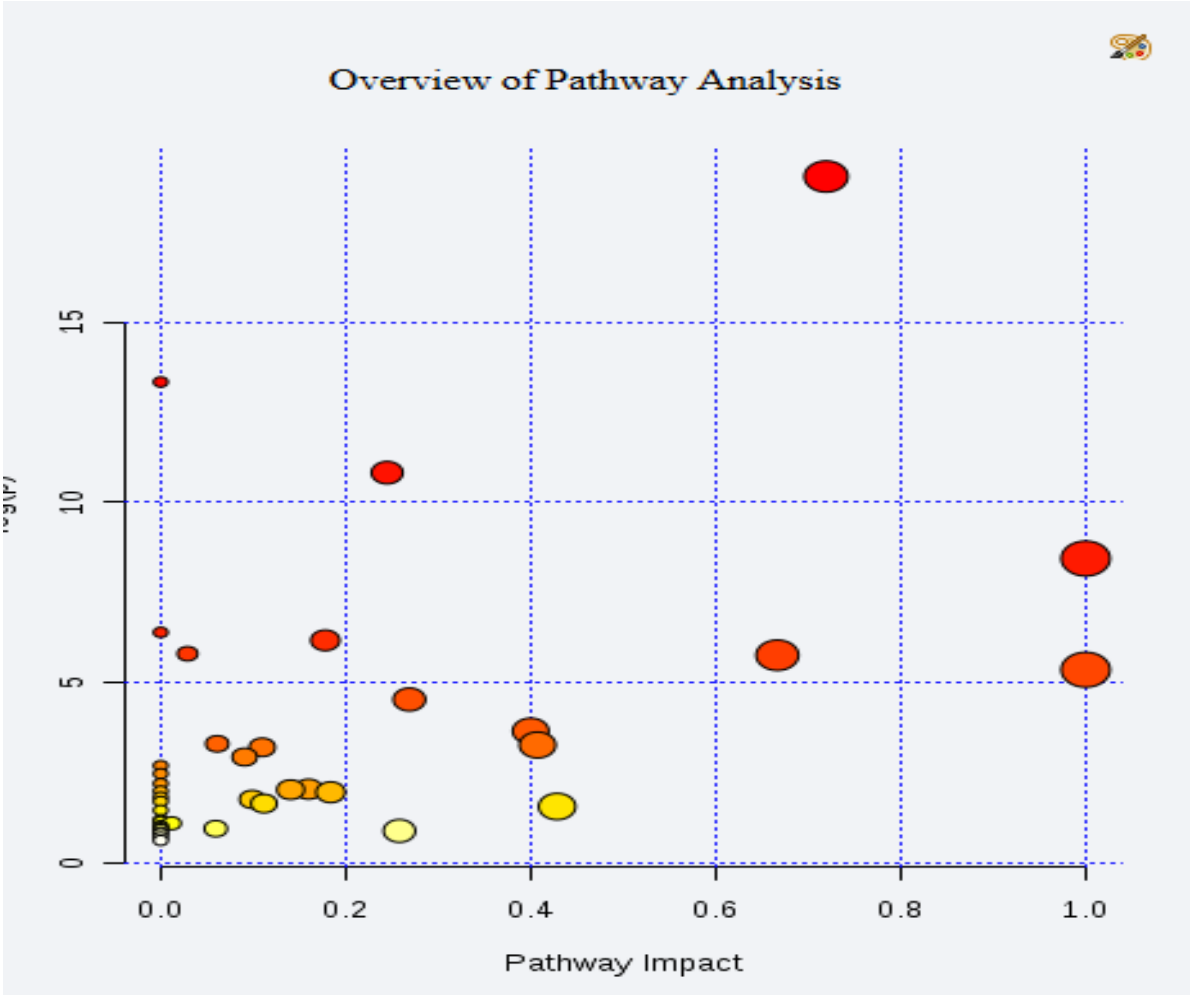

Figure 20. Pathway overview showing from top to bottom on the graph in descend order of importance the different pathway involved in this NPC mouse model liver dysfunction disease study. Thus, the significance of the pathway decreases going from red to yellow. Therefore, alanine, aspartate, and glutamate metabolism (–log(p) of 19.032), followed by aminoacyl-tRNA biosynthesis (–log(p) of 13.335), then arginine and proline metabolism (–log(p) of 10.82) and finally glutamine and glutamate metabolism (–log(p) of 8.4389), appeared as the most significant pathways in this disease dataset.

- **Pathway Analysis of Alanine, Aspartate and Glutamate Metabolism**

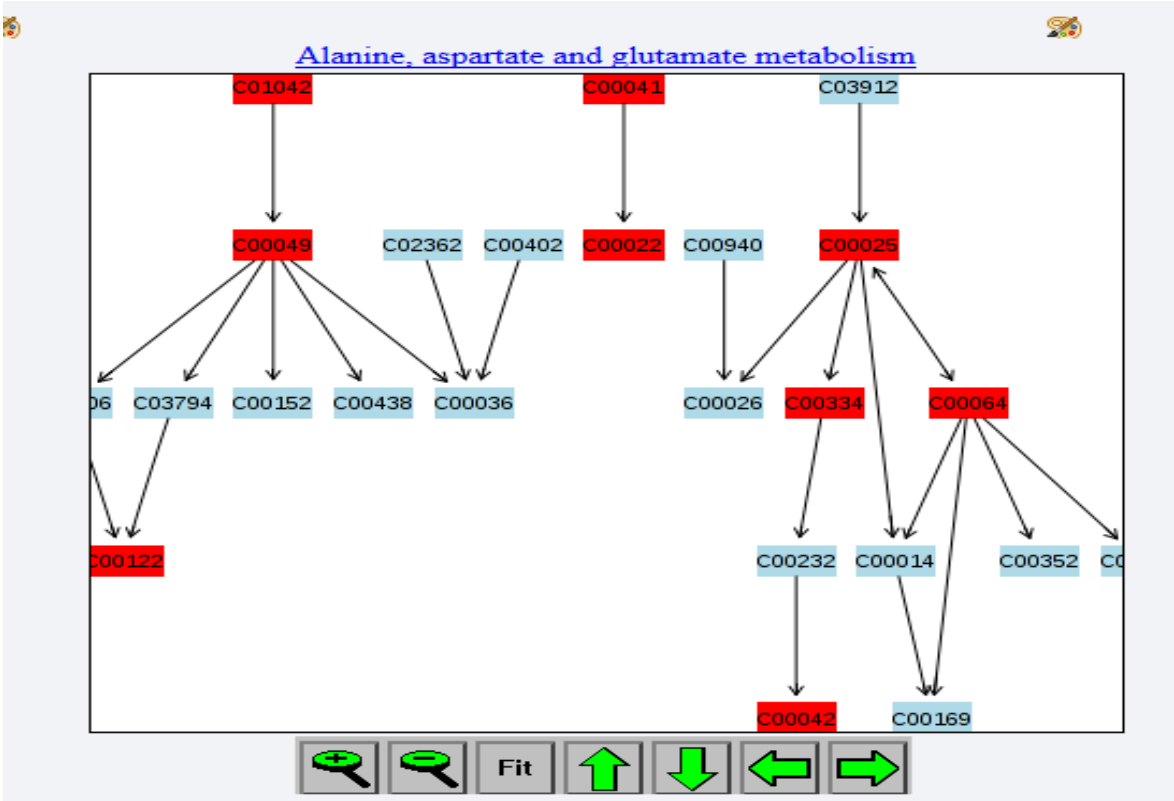

Figure 21. Pathway analysis **alanine, aspartate and glutamate** metabolism, where the following statistics are presented; Holm adjusted p values of 4.4484E-7, a FDR of 4.4484E-7 and a high pathway impact of 0.7194. Hence, this pathway is highly differentially expressed in the NPC (mice model) liver disease.

The current pathway is highly differentially expressed, and involves, the N-acetyl-L-aspartate (C01042), which can generate L-aspartate (C00049), which in turn can be transformed to fumarate through different intermediates metabolites not present in this dataset. However, L-alanine (C00041) can produce pyruvate (C00022), and also, L-glutamate (C00025) can generate L-glutamine (C00064) directly, or gamma-aminobutyrate (C00334), which in turn, produces succinate (C00042). These metabolites featured highly on this pathway except gamma-aminobutyrate which is not present on the dataset.

- **Pathway Analysis of Aminoacyl-tRNA biosynthesis**

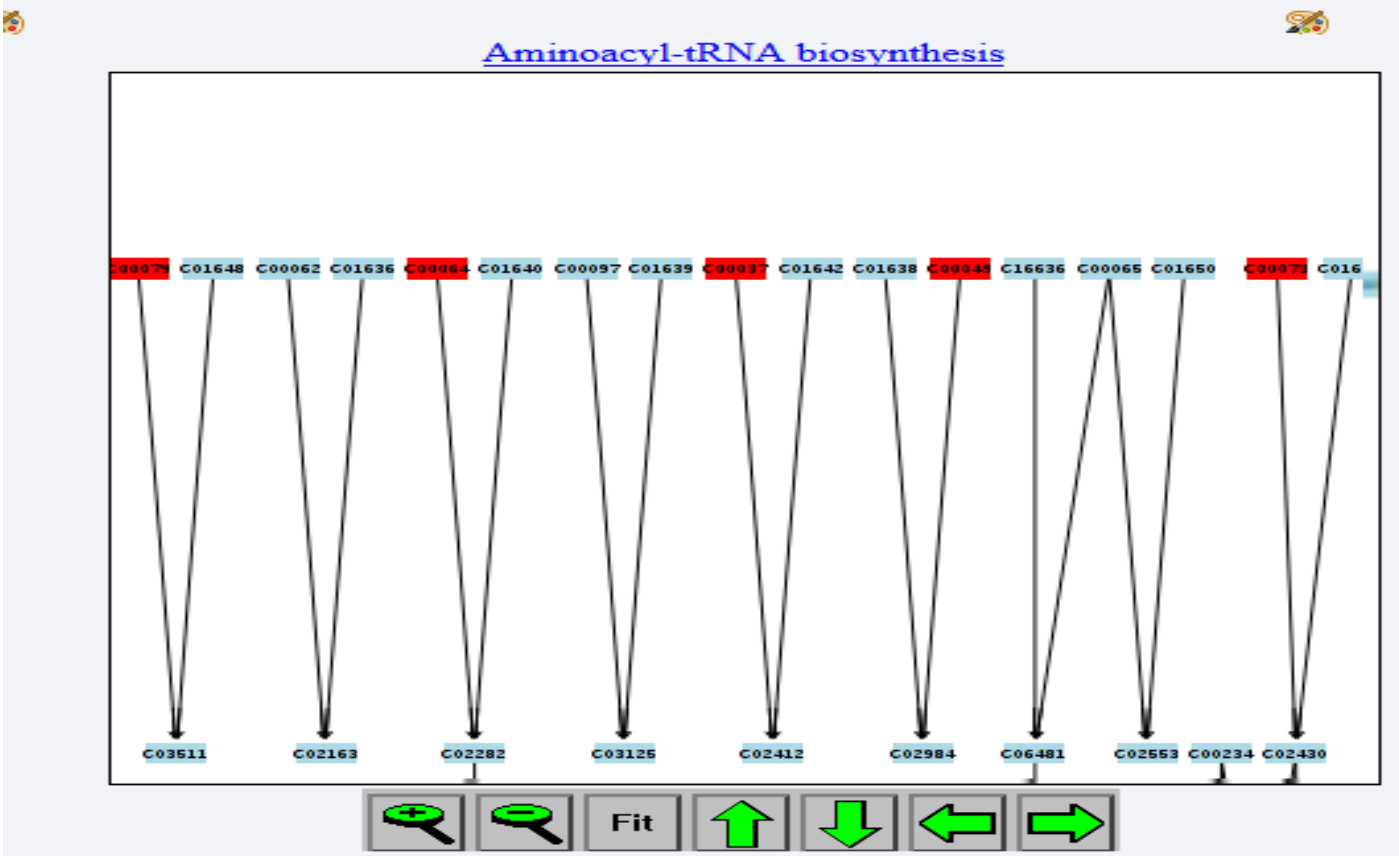

Figure 22. Aminoacyl-tRNA metabolomics pathway analysis with a Holm adjusted p value of 1.3096E-4, a FDR of 6.6287E-5, and a pathway impact of 0.0, and the second highest -log(p) of 13.335). Therefore, this pathway is differentially-expressed, with major contribution of L-glutamate (C00025), L-tyrosine (C00082), L-leucine (C000123), L-isoleucine (C00407), L-lysine (C0047), L-alanine (C00079), L-methionine (C00073), L-aspartate (C00049), glycine (C00037), glutamine (C00064), and L-phenylalanine (C00079).

Several metabolites were observed in the present ^{1}H NMR spectra and are monitored with regard to aminoacyl-tRNA biosynthesis. The current pathway analysis could start from the L-glutamate, then to L-tyrosine through intermediaries which can also generate L-leucine through intermediaries. The latter can produce L-isoleucine, leading to L-lysine, then L-alanine, leading to L-methionine, then L-aspartate, producing glycine, then glutamine and finally, L-phenyalanine are all highly featured on this pathway. However, this series of reactions are performed using intermediaries most of which are not present in this dataset and are excluded from the enrichment analysis. It is important to mention that glutamate an anion of glutamic acid is used by nerve cells to send signals to other cells.

- **Pathway Analysis for arginine and proline metabolism**

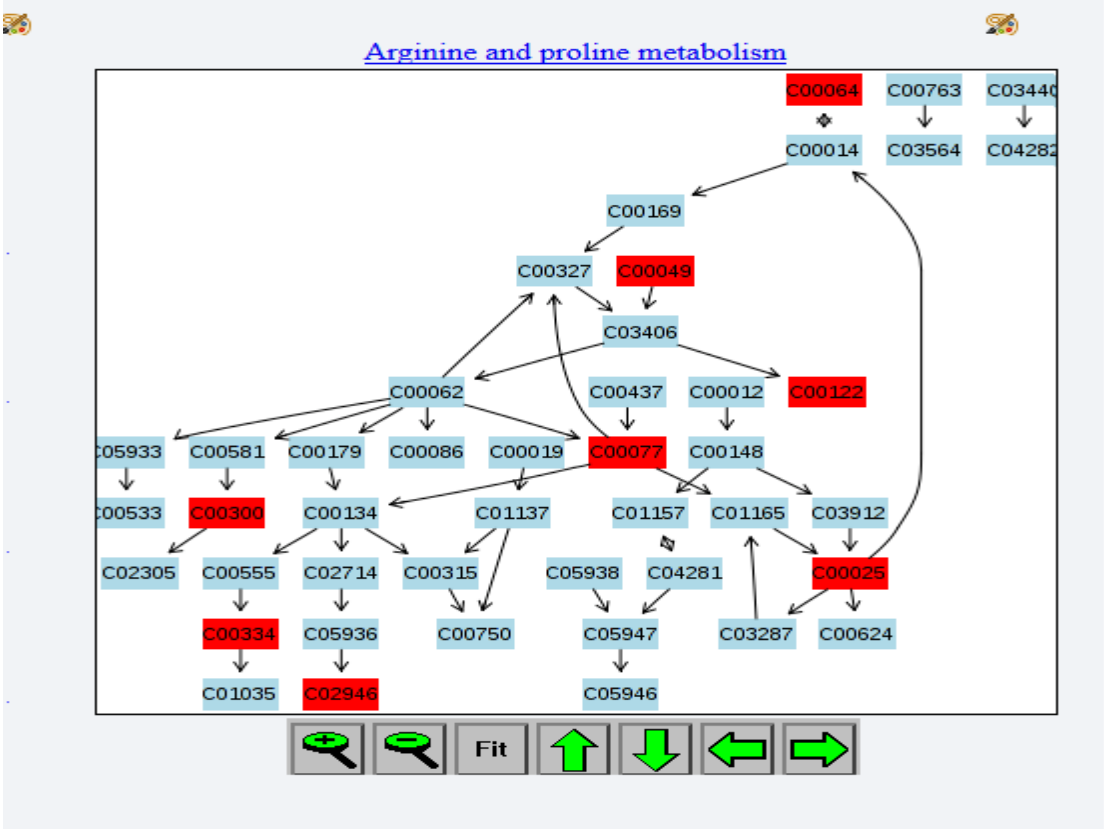

Figure 23. Arginine and proline metabolomics pathway analysis with a Holm adjusted p value of 0.0015996, a FDR of 5.4652E-5, and a pathway impact of 0.24479. Therefore, this pathway is strongly differentially-expressed, and features modified concentration of p-hydroxyphenylacetate, **hippurate, pyruvate**, fumarate and succinate.

L-glutamine (C0064) generates the ammonia (C0014), which in turn can go through intermediaries to produce L-arginine. Indeed, the L-arginine (C0062) can be generated going through two intermediaries, which in turn will produce **creatine** (C00300). On the other hand, arginine (C0062) can produce **ornithine** (C00077) which can generate **glutamate** (C0025). However, glutamate can generate also ammonia. Nevertheless, peptide (C0012) can produce proline (C00148), which in turn can be converted to pyrroline-5-carboxylate (C03912) which produces L-glutamate. All the biomolecule highlighted in bold are highly present on this pathway. Nevertheless, in the liver, especially in the cells surrounding its central membrane, ornithine is used for the synthesis of glutamate and glutamine, and can also be used in peripheral tissues for the synthesis of both glutamate and proline (Cox and Nelson, 2013).

Dynamic network of cholesterol trafficking plays the important role of maintaining the cholesterol content of subcellular membranes. For instance late endosomal transmembrane NPC proteins in the lysosome lumen interact for the transportation of the cholesterol from the endosome to the plasma membrane and the endoplasmic reticulum. The unesterified cholesterol will be accumulated at endosomal level as a consequence of a dysfunction (Kennedy et al., 2016).

- **Glutamine and Glutamate Metabolism**

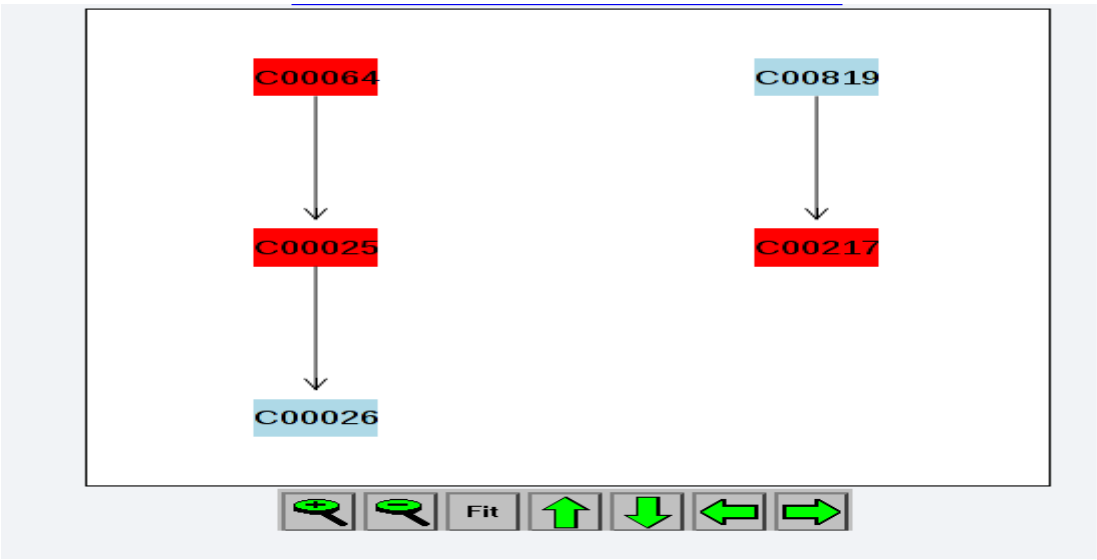

Figure 24. Pathway analysis of glutamine and glutamate metabolism, with the following statistics; Holm adjusted p values of 0.017087, a FDR of 0.004434 and a high pathway impact of 1.0. Hence, this pathway is highly differentially expressed in the NPC (mice model) liver disease.

The current pathway is highly differentially expressed, and involves, L-glutamine (C00064), which can generate L-glutamate (C00025), and the latter can be converted to oxoglutarate (C00026) absent in this dataset. Glutamine can directly produce glutamate, and the metabolism of the latter involves several reversible or irreversible reactions. Thus, the transformation of the L-glutamate to N-acetyl-L-glutamate takes place in the presence of N-acetylglutamate synthase (NAGS) with the help of L-arginine as activator, and several inhibitors, including succinate, coenzyme A, N-acetyl-L-aspartate and N-acetyl-L-glutamate (SHIGESADA and TATIBANA, 1978; Yelamanchi et al., 2016). In the same manner, the conversion of glutamine to glutamate is a reaction catalysed by a number of enzymes such as glutaminase (GLS/GLS2),

glutamine-fructose-6-phosphate transaminase (GFPT1 and GFPT2) and finally phosphoribosyl pyrophosphate amidotransferase (PPAT) (Hu et al., 2010; Yelamanchi et al., 2016). It should be noted that glutamate is the highest concentration of molecule in the brain cells and is the canal used by the nerve cells to send signal to other cells.

- **General Consideration:**

^{1}H NMR spectra generated from aqueous extracts of mice liver samples with a group of the NPC liver dysfunction disease (NPC LDD) carriers and a control group representing the healthy mice were very informative. Amongst important biomarkers features amino acid such as alanine, leucine, isoleucine, valine etc.; organic acid anion such as glutamate, formate, fumarate, lactate, and pyruvate. Other resonance features in these spectra were assignable to methylalanine, dimethylamine, etc. Moreover, bile acids which are usually associated with taurine or glycine in the liver were also identified by the different modelling techniques employed, including SPC, ESPC and Heuristic. The scalar visualisation technique could detect these major biomarkers from the raw data analysed without any strong data pre-processing technique applied to it. Moreover, the OSVM could unambiguously segregate between the two classes of liver extract samples with a high-performance value from the AUC ROC that was selected as better performance metric than the Gmean. Finally, using the PCR model, allowed the selection of major biomarkers, together with their pathway analysis that connected a different group of metabolites and the underlying development of the NPC (mice-model) liver dysfunction disease aetiology.

Following the detection of these biomarkers and the analysis of the main pathways related to them, the next step is dealing with the modelling techniques validation. This should give more credit to the TRT model developed together with the findings.

5.5. Research Model Validation

5.5.1. Validation with the Intelligent Tri-Modelling Techniques (ITMTs)

The intelligent tri-modelling techniques implemented in this thesis use three different algorithms for modelling the different datasets involved in this research. This encompasses the use of three modelling techniques. The first algorithm used is the scalar visualisation algorithm (SVA). This is a valid technique acquired from the Matlab toolbox and creates a 'look-up' table of colours that are matched with indexes based on the resonance of the ^{1}H NMR spectra of the different metabolites found in the datasets. The algorithm is valid and has been already used in different research studies to gain insight into data structure and the relationship between features (Johnson, 2015; Johnsona, 2012; Komura, 2016; Peng, 2011).

The second model used is the optimum support vector machine (OSVM). This model is a support vector machine tuned to its optimum performance using the maximum 'tuning' capability of the different parameters used in the SVM. Similar SVM has already been developed, and this includes those focused in improving classification on imbalanced datasets, or developing efficient models (Köknar-Tezel and Latecki, 2011; Maji et al., 2013).

The third algorithm used in this research is the principal component regression (PCR). Although the current PCR involved the PCA algorithm for the detection and selection of principal components PCs, and the multiple logistic regression (MLR) strategy allowing to regress the dependent variable Y on the PCs. The same technique has been additionally used in several research studies, and has been well validated (Boston University School of Public Health, 2013; Kumar and Madhu, 2012).

Another level of validation was implemented based on the research results obtained.

5.5.2. Research Result Validation

Different techniques were implemented in order to validate the results obtained. This includes the stratified k folds cross-validation technique enabling to generate k subsets of the dataset with k-1 subsets used for training and 1 subset for testing. This helps us to assess how the system learns from training processes. However, the stratification technique allows to have in each subset a fair representation of the minority group. In addition, the ADASYN technique that creates synthetic samples in the minority class in order to balance the dataset and alleviate the effect of bias in favour of the majority class was also applied. Moreover, variables related

to the support vector machine were ‘tuned’ based on the training and testing sets in order to permit the classifier to perform at its maximum performance level, and hence the label of optimum support vector machine (OSVM) for the SVM employed in this thesis.

It should also be noted that the results of this research are validated against the appropriate standard, based on the knowledge that the different techniques used are all standard and valid techniques in data classification. This includes k-fold cross-validation techniques, grid search methods, the ADASYN technique, the variable selection techniques, etc. In addition, the different classification performance rates based on the AUC ROC and Gmean values that are means of the different models were run several times to generate conclusive results, i.e. AUC ROC = 0.944 and Gmean = 0.885 as mean values for NPC1 disease dataset. Furthermore, SVA could detect several major biomarkers identified by all three algorithms involved in the present intelligent tri-modelling techniques. For example, SVA could highlight pyruvate, glutamate, lactate, 2-hydroxyisobutyrate, 3-hydroxybutyrate, L-fucose, 3-hydroxyisovalerate, 2-hydroxyglutarate, etc., as major biomarkers for NPC1 disease diagnosis, and valine and leucine in the NPC LDD dataset. Finally, the result produced in relation to the validation of the biomarkers discovered is discussed below.

5.5.3. Independent statistical validation

Effective validation of the biomarkers discovered should involve a combination of a clinical and statistical processes. The clinical validation will not be discussed in this research because it is beyond the scope of it. However, the independent statistical validation process that intends to establish a correlation between miglustat treatment and these biomarkers is succinctly highlighted in this sub-section.

Biomarkers discovered required candidate biomarkers to undergo two stages in the validation process, including clinical and statistical validations. The statistical validation, which has been covered in this thesis, included an initial stage where a correlation will be established between the biomarker(s) discovered and NPC1 disease diagnosis. This stage has been completed with the discovery of the potential biomarkers using NPC1 disease patients’ blood plasma and liver extracts of mouse model datasets. Secondly, independent statistical testings will help us to determine correlations between the miglustat treatment and any improvements observed in

NPC1 disease patients' conditions (Buyse et al., 2010). The intelligent algorithm development (IAD) presented above (see 3.2.2) highlights stages of an independent validation process for such biomarkers. Indeed, if miglustat treatment reduces the concentration of the biomarkers discovered, which include hippurate and adipate in blood plasma, then these biomarkers may indeed be considered as valid for this NPC1 disease study.

Finally, the processes implemented during this research were validated against the theoretical framework developed (ITTFM) noted below.

5.5.4. Research theory validation

The theory of intelligent technology task fit model (ITTFM) has to be validated against standard. The diverse set of operations taking place during experimental manipulations, including the disease samples selection, the NMR process and spectra quality, the SVA processing time, the SVA major biomarkers detection processes, and accuracy, the OSVM classification performance rates, the PCR biomarker detection process, etc., could be considered as ideal characteristics of the measurement of the performances and the validation of the ITTFM (Albro et al., 2010; Bird et al., 2014; Chau et al., 2008; Murphy and Kerr, 2004; Venkatraman, 1988).

Therefore, the possibility of using the different activities taking place, in order to ensure that they have been verified, as a mean of validating the model's suitability, could be undertaken too. This technique could be regarded as a validation step of the ITMTs. Similar studies have been conducted to validate the theory developed (Albro et al., 2010; Venkatraman, 1988).
In the present thesis, the theory of the ITTFM related to the human involvement in the process of detection, discovery of biomarkers in disease diagnosis. Therefore, the concept of 'fit-as-match' attempts to detect the level of matching between the process and the high standards required in biomarker discovery. However, the concept of 'fit-as-deviation' investigates the level of deviation between the process and the algorithm employed for biomarker discovery; while the 'fit-as-profile' concept determines the level of convergence between the process and the profile defined; both these processes have been investigated.

- **Consideration**

The model profile is related to the different variables necessary in the current biomarker discovery process. They include the two main steps that are data collection and data analysis. Setting process scheduler allows us to ensure for example that different batch of the blood samples going through NMR spectroscopy are run identically for consistence, repeatability, reproducible, and verifiable. This was implemented during data collection, data analysis, and is now used as technique and standard in the validation of the intelligent technology task fit model.

- **ITTFM model validation**

Processes	Fit-as-profile	Fit-as-match	Fit-as-deviation	Comments
Data collection	Define a metric for data collection	Matching normal process	No deviation observed	Standard was met for all Data Collection stages
Raw samples preparation and processing	Define a metric for samples preparation	Matching normal process	No deviation observed	Samples prepare according to standard
NMR samples processing - NMR spectrometer	Define a metric for NMR data collection	Matching normal process	Minimum deviation due spectrometer failure (technical issues)	Standard was met for NMR spectroscopy– Spectra quality, Peaks Alignment, Baseline Correction
Data Processing - 1D NMR Processor	Define a metric for data processing	Matching normal process	No deviation observed	NMR data was processed according to
Features selection	Define a metric for features selection	Matching normal process	No deviation observed	Features selection standard was achieved ADASYN – NNFS – NNO - etc.

Grid Search Technique	Define a metric for SVM parameters selection	Matching normal process	No deviation observed	Standard was met with regards to parameters selection and tuning
SVA for Data Visualisation	Define a metric for data visualisation	Matching normal process	No deviation observed	Visualisation Standard was met for all the datasets.
SVM for Data Classification and Prediction	Define a metric for data classification	Matching normal process	No deviation observed	High classification Standard was met for all the datasets
PCR for Biomarkers Discovery	Define a metric for biomarker discovery using PCR	Matching normal process	No deviation observed	High biomarkers discovery standard. Detection of known and probable new biomarkers.

Table 46. Matching with standards in data collection, data processing and analysis research model validation using the ITTFM theory. The different characteristics defined in alignment levels models (Table 3) allow us to validate the ITTFM model.

The intelligent technology task fit model was validated against standard, especially against the theory of the task technology fit model developed by Goodhue and Thompson by applying rigorous scientific processes that are genuine, sequential, replicable, and reproducible. In addition, the different steps involved can be assessed throughout the development process (Albro et al., 2010; Bird et al., 2014; European Commission, 2010; Goodhue and Thompson, 1995).

5.6. Chapter Summary

The data analysis and results chapter include an introductory section that highlighted the main sub-sections of the chapter. Subsequently, the intelligent tri-modelling techniques (ITMTs) for

the NPC1 disease (human model-based) data multivariate analysis, using the scalar visualisation algorithm for data visualisation, the optimum support vector machine for data classification and the principal components regression for data classification, regression and biomarkers discovery was explored. The ITMTs were then explored for the liver dysfunction disease dataset (mouse model-based) multivariate analysis. Hence, these represent the three main algorithms used for the scrutiny of the analysis of the dataset to seek important biomarkers. Pathway analysis of these biomarkers was then carried out, allowing us to have more insight into their relation and involvement in the NPC1 disease diagnosis in general. Following this approach, the research model was validated, including validation of the ITMTs, and the results of the research performed within it. Finally, the data analysis and results chapter in general was concluded with a brief summary section.

"*Logic will get you from A to B. Imagination will take you everywhere*"

Albert Einstein.

6. DISCUSSION

6.1. Introduction

In the precedent chapter, data analysis and the results chapter explored different processes such as operating different algorithms and analysing the results obtained. In this respect, different biomarkers were detected amongst those already established such as glutamate, pyruvate, taurine, hypotaurine, etc. in the NPC1 disease diagnosis set, and the phenylalanine, glutamine, trigonelline, pyruvate, fumarate, etc., in NPC1 liver dysfunction. Others are less well-known as potential biomarkers in the NPC1 liver dysfunction disease diagnosis (mouse model-based), and these include adipate, xanthine, hippurate, etc.

In this chapter, the focus is placed on discussing results produced and the different implications with regard to our understanding of the underlying transformations at the molecular level, especially the related pathways analysis involved. In this manner, two main discussions will take place in this chapter. In the first section, the discussion will involve the NPC1 disease diagnosis, with related biomarkers discovered and the relevant pathway analysis. In the second section, the discussion is related to the NPC1 liver dysfunction, and the relevant pathway analysis to consider. In the final section a summary of the discussion section will be presented.

6.2. Discussion Related to the NPC1 Disease Diagnosis

6.2.1. Discussion Based on the NPC1 Disease Statistical Analysis

The statistical analysis performed was built on the following logic, which includes starting by segregating between the diseased group and the wild treated/healthy control one. The optimum support vector machine (OSVM) is used as a deterministic classifier to establish separation. Henceforth, a visualisation algorithm is used to understand the underlying relationships between the disease features responsible for this clear separation. Finally, the principal component regression (PCR) is employed as a probabilistic classifier to predict the influence of disease features. However, because the current PCR employed the principal components analysis (PCA) and then multiple logistic regression (MLR) analysis of PC scores vectors, it allowed the dimensionality reduction, the visualisation and ascertainment of the two classes, the former defined by the PCs, and finally ranking the features in terms of their ability to predict their effects on disease progression.

The optimum support vector machine was able to correctly distinguish between the class of NPC1 disease patients and the wild-type (WT) healthy individuals with a higher level of performance.
In this regard, it uses two main functions and two main performance metrics that are for the functions the radial based function (RBF) and the linear function in the data classification fields, and includes the geometric mean (Gmean), and the area under the curve receiver operating characteristics (AUC ROC) for the performance metrics. A combination of the RBF and AUC ROC approaches was found to yield improved performances compared to the others. For example, the Gmean average performance based on the RBF function was 0.88530 (Table 5), while the Gmean average performance based on the linear function was 0.871 (Table 6). The AUC ROC average performance based on the linear AUC ROC function was 0.937 (Table 7), while the AUC ROC average performance based on the RBF function 0.944 (Table 8). These were expected results in line with different research carried out using the same function and performance metrics (Martí and Reinelt, 2011; Tang et al., 2009). Indeed, linear function was observed to be less reliable than the RBF function, furthermore, the Gmean approach was less effective than the AUC ROC one.

Consequently, most of the results discussed are based on the RBF - AUC ROC combination applied. Hence, the OSVM participants' discriminative power successfully differentiates between the NPC1 diseased patients and the healthy participants. A thorough analysis was

carried out, and different parameters were used to assess their influence on classifier performances. The current research has also shown that increasing the number of k folds in the cross-validation, the number of times that the stratified cross-validation is repeated, the number of nearest neighbours in the features selection, the proportions of features used, and finally increases in oversampling improve the OSVM model's overall performances.

This can be explained by the fact that increasing the number of k folds increases the number and size of the training set whilst reducing the corresponding number and size of testing set. Hence, the OSVM is trained on a bigger range of data and testing on a smaller set. Furthermore, the computational time of the classifier increases. Likewise, increasing the number of times the stratified cross-validation training is repeated allows the classifier to 'learn' more from the dataset. However, there is a risk in this case of 'overfitting' the 'learner' that might indeed learn the dataset too well to be able to perform on unknown dataset (Buduma, 2015). Nevertheless, the stratified cross-validation technique is able to tackle and avoid this potential risk of overfitting. Increases in the number of k folds and the number of repetitions has a limit over which the classifier performance is affected by constant decreases. For example, the number of k folds varied from 5 - 15. However, the number of repeats were kept under a more reasonable proportion in the range of 10 - 20, especially since increases in the processing time, and the high level of processing effort required.

Results achieved by the OSVM in terms of feature classification in two major classes is further support by the fact that a clear visualisation difference could be made between these using biplots in the principal components analysis PCA (Figure 8).
Nonetheless, the focus in this discussion section is on the scalar visualisation algorithm. The statistical results obtained from the NPC1 disease dataset presents significant changes on the ^{1}H NMR profile of the NPC1 disease patients' class compared to that of wild treated class (Figure 6).

Henceforth, the scalar visualisation algorithm (SVA) displayed a significant change in the following areas. The first area covering the chemical shift regions 0.81-0.89 ppm corresponding to the metabolites ***hexacosanoate***, the area covering the chemical shift 0.89-0.95 ppm corresponding to the biomolecule ***L-isoleucine***, and the area 0.95-1.06 ppm corresponding to both ***L-Leucine*** and ***L-Valine.*** These three branch chain amino acids (BCAAs) encompassing valine, leucine and isoleucine are very important biomarkers in the

aminoacyl-tRNA biosynthesis pathway, because they can initiate the formation of aminoacyl-tRNA derivatives, which can generate in turn proteins production. However, excessive accumulation of protein can lead to chronic liver disease such as liver fibrosis and liver failure (Bataller and Brenner, 2005; Monirujjaman and Ferdouse, 2014). Indeed, this selection is confirmed by the pathway statistics including the very low values of the Holm adjusted p values and the FDR value < 0.05. Additionally, the presence of the valine, leucine, isoleucine **and pyruvate** have been detected by the PCR, which ranked them at a higher level, and this will be further discussed below.

The second group of biomarkers detected by the SVA are related to the chemical shift region 1.21-1.31 ppm corresponding to the ***(R)-3-hydroxybutyric acid***, and ***L-fucose***; the chemical shift range 1.31-1.37 ppm corresponds to the ***lactate***, ***3-hydroxyisovaleric acid*** and the chemical shift 1.45-1.50 ppm (less pronounced) corresponds to L-alanine, which has been shown to be important in acute liver failure (Thodou et al., 2017). On the other hand, pyruvate is very important in aminoacyl-tRNA biosynthesis performed in liver, where it is transformed into glucose using the adenosine triphosphate ATP, a very high energy metabolite. Moreover, lactate, succinate, etc., can be transformed into glucose. This reaction is reversible, with glucose being transformed during glycolysis into pyruvate using ATP, and then to lactate (Silva, 2002). However, glucose can be transformed into fatty acids and stored at the lysosomal and endosomal levels, with the risk of aggravation of NPC1 disease.

Furthermore, plasma filled contour plot could detect important change of structure in the 0.81-0.89 ppm chemical shift region corresponding to ***hexacosanoate or cerotate***. These molecules are long-chain saturated fatty acids, important for the build-up of fat in the human body. These three sets of biomolecules are the most important biomarkers for the NPC1 disease diagnosis by the SVA technique.

The third group of biomarkers detected by the SVA strategy, although to a lesser degree, include the chemical shift 1.96-1.98 ppm, corresponding to the biomolecule ***methylglutarate*** and that covered by the bucket 1.98-2.03 ppm chemical shift range, corresponding to ***2-hydroxyglutarate***. Other chemical shift buckets detected are 2.12-2.17 ppm corresponding to ***L-methionine, and L-glutamine,*** and the 2.34-2.39 ppm one for ***pyruvate and glutamate***. Moreover, the chemical shift region 2.68-2.74 ppm corresponds to the Citrate, which is a tricarboxylic acid anion important in the metabolism of aerobic organisms, while the 2.85-

2.88 ppm zone corresponding to trimethylamine; these compounds were detected in the plasma filled contour plot and the full boxplot with a medium peak amplitude. The SVA modelling technique suggested only a medium contribution in the explanation of the NPC1 disease difference observed.

The visualisation algorithm (SVA) using the principal of look-up table correlates the ^{1}H NMR bucket values as an index to a unique colour located in the Table. Therefore, the difference in colour ranging from blue (0) to red (0.35) shows the extent of the transformation occurring at molecular level. Indeed, the level of transformation is shown on the features maps with the variation in colour (Figures 6 & 7) and the boxplot with amplitude of the boxes (Figure 5). This shows that the features number 13 and 14 correspond respectively to the chemical shifts buckets 1.21-1.31 and 1.31-1.37 ppm corresponding to lactate and 3-hydroxyisovalerate, and therefore they can be considered as major biomarkers in the current disease aetiology with respect to SVA modelling approaches.

The SVA uses a scalar algorithm to map the underlying changes at the molecular level, and displays them on the colour map representing different biomolecules, together with the relationships existing between the diseases' metabolome. As a consequence, molecules and their particular characteristics can be visualised, in such a manner that allows major biomarkers to stand out from the tens of different biomarkers detected in the multivariate analysis (Figure 6 & 7). For example, the plasma contour plot (Figure 6) highlights the pattern differences and changes in the feature map, with change in colour and brightness showing a difference between features in columns 6, 13, 14, 25, 26, and 27. These columns correspond, respectively to the chemical shifts buckets 0.81-0.89, 1.21-1.31, 1.31-1.37, 2.03-2.09, 2.09-2.12, and 2.12-2.17, which in turn are the hexacosanoate, the (R)-3-hydroxybutyrate and L-fucose; the lactate; the 3-hydroxyisovalerate; the Citrate; the N-acetyl-4-O-acetylneuraminate; methionine and glutamine. For example, the glutamine's importance has been highlighted in the aminoacyl-tRNA biosynthesis pathway (see 5.3.5). However, methionine is used for methyl group transfer in most biosynthetic reactions, especially in amino acid catabolism. The methyl group transfer is preferably accomplished by the S-Adenosylmethionine (adoMet) used as cofactor. The adoMet is produced by the action of the methionine and ATP, catalysed by methionine adenosyl transferase. This reaction, in which the sulphur atom is involved in a nucleophilic substitution by attacking the methyl group of the ribose moiety rather than the carbon on the other phosphorus atoms. Indeed, the sulphur as an electron rich nucleophile quickly attack the

methyl group of the ribose, which is partially positive charged. This in turn confirms the importance of methionine in the NPC1 disease (Cox and Nelson, 2013; Iyer and Hengge, 2008).

It should also be noted that most of these biomarkers were also detected by other modelling techniques employed in this research, especially that involving principal component regression PCR. This data analysis technique was employed for a probabilistic classification that ultimately could determine and detect major biomarkers using the intelligent tri-modelling techniques (ITMTs) (Schöfl et al., 2016). The PCR strategy firstly employed principal component analysis scores vectors and then multiple logistics regression (MLR). Their respective statistics are highlighted with their importance in the features ranking (see 5.4.4).

Different PCs and their corresponding r values obtained by running MLR several times, were used in combination with the correlation factor between the original features and the PCs. Table 11 above gives the main PCs involved in the 90%, variability together with the deviance of the fit (67.98) and the estimated dispersion parameter (0.832 < 1). Based on these different statistics and the level of the ranking achieved, 80% and 90% variability were considered the most reliable. Other statistics have to be considered; for example, although 80% variability model has larger deviance of fit (93.92 > 67.98) and estimated dispersion parameter, gave rise to an improved level of ranking, and appeared to be more reliable compared to 90%. In addition, 80% variability shows more consistency in the ranking of the 10 main biomarkers out of the 24 ranked using all the four levels of variability, i.e. 90, 80, 70 and 60% variabilities; with the correct ranking (CR) following values 95.83% for 80% variability, against 91.67% for 90% variability models (see 5.4.4). In this manner, the 80% model was finally chosen and confirmed as providing the most effective level of variability.

This selection/ranking of the level of variability is based on a certain level of logic. Indeed, ranking the biomarkers is one of the main objectives of biomarker discovery. Knowing which ones carry the maximum weight for disease diagnosis purposes is of much greater importance. For example, the biomarkers importance can be related to the impact value of the pathways implicated for NPC1 disease. However, the 4 levels of variability were assessed based on this same NPC1 disease dataset. Therefore, the disease factor is a constant related to all the 4 levels of variability, and use of the consistencies in rankings, and the level of correct ranking (CR) is a valid means of choosing the most valuable level of variability.

Another key point is that the 80% and 90% levels of variability could detect with differing levels of importance, three key biomarkers amongst the 24 top ones selected. These include pyruvate, and glutamate, isoleucine, and also glutamine. However, the 70% and 60% levels of variability could detect only the first three biomarkers, including pyruvate, glutamate, and the isoleucine. The importance of glutamine and isoleucine in the aminoacyl-tRNA biosynthesis pathway has been highlighted, and will be further discussed within the next sub-section (6.2.2) related to pathway analysis. Pyruvate is also important in phenylalanine metabolism and in glycolysis. Through glycolysis, glucose is transformed into pyruvate, and therefore pyruvate plays a key role and also maintain blood glucose levels via gluconeogenesis (Godoy et al., 2010; Silva, 2002). Notably, pyruvate was detected as one of the main biomarkers amongst the 7 main ones selected by this probabilistic modelling technique.

Other very important techniques used to support the different level of variability in ranking the different features were the tri-raking techniques (TRT). Indeed, TRT played a main role in feature rankings, and their performance level is therefore dissected and discussed. This includes the sum of the products of the coefficients (SPC), the exponential sum of the product of the coefficients (ESPC), and finally the heuristic model. Two of these have had the same level of performance throughout and include the SPC and the ESPC approaches. The heuristic model being the only one performing differently; hence, it is used to assess the differences between the performances.

The SPC model was used as a reference, since the development of the other two feature ranking techniques were based on the mathematical fundamental relating the coefficient of regression r and the coefficient of correlation α to the contribution of each feature to the regression line. Thus, these ranking techniques were introduced based on the SPC technique. It is therefore, a question of logic that SPC be used as a reference technique for the assessment of the other two models.

The TRT model developed achieved ranking scores of 100% for ESPC at all four levels of variability, while heuristics achieved the following; (for 90% variability) CR values of 91.67%, (for 80% variability) 95.83%, and (for 70% and 60% CR variability) 87.50%. These results place SPC and the ESPC as the most effective ranking methods, and with the heuristic technique being less effective. This classification or ranking of the three techniques is

consistent with the logic associated with their development. In the light of this ranking, biomolecules primarily detected by these techniques are provided below. In a selection of the 10 best out of a selection of 24, the tri-ranking techniques identified adipate, pyruvate, glutamate, lactate, 2-hydroxyisobutyrate, 3-D-hydrodybutyrate, L-fucose, 3-hydroxyisovalerate, 1-methylhistidine, quinolinate, methylxanthine, and 4-hydroxybenzoate. From this first selection of the 10 best markers in the NPC1 disease development were detected; including the pyruvate, the glutamate, and the lactate.

Isoleucine and pyruvate were ranked respectively 13th and 3rd best biomarkers respectively in understanding this disease's underlying biochemical effects, and hence the importance of isoleucine in aminoacyl-tRNA biosynthesis, and that of pyruvate in the alanine, aspartate, and glutamate metabolism. Moreover, all the ranking techniques performed similarly with different ranking abilities, and expressing different levels of importance based on the technique. Hence, the sum product of the coefficients (SPC) and the exponential of the sum product of the coefficients (ESPC) have ranked the isoleucine in the 17th position, and valine and leucine at the 40th place.

However, the heuristic approach ranked them the 26th and 42nd best biomarkers for NPC1 disease diagnosis. Additionally, this heuristic technique failed to rank the isoleucine amongst the 24 best features, unlike the other two techniques applied. This can be explained by the fact that the heuristic search technique, by definition, is one which allows to find a more rapid solution which may not be the best; hence, a 'trade-off' has to be made between accuracy and the speed of the analysis (Masulli et al., 2013; Shi and Ólafsson, 2008).

Nonetheless, when extended to the 24 best markers, more of these were captured in this ranking/selection process. For example, glucose, glutamine, methionine, etc., were included in the selection, which are also important in the different pathways detected. During the glycolysis, the glucose is broken down in the cytosol, while pyruvate the product of glycolysis is transformed to Acetyl CoA. On the other hand, through gluconeogenesis the glucose is transformed to pyruvic or fat, with the former being also an important intermediate in fat production (Cox and Nelson, 2013). Whereas, glutamine, methionine and histidine are shown to be important in aminoacyl-tRNA biosynthesis, glutamine can generate ornithine which, in turn, can be transformed to pyruvate. The latter importance has been already established (Pubchem, 2017; Ruiz-Rodado et al., 2014).

6.2.2. Pathway Analysis and NPC1 Disease Diagnosis

With regard to the NPC1 disease diagnosis, the main pathways and related biomarkers involved will be scrutinised through an extensive discussion in order to permit a thorough understanding of disease diagnosis and prognosis. The NPC1 disease dataset pathway analysis and ranking in accordance to their importance in the disease aetiology, are based on the Holm adjusted p value to select the following pathways as the main ones involved in the disease diagnosis and progression. Three main pathways detected are outlined below.

- **Aminoacyl-tRNA Biosynthesis**

Aminoacyl-tRNA is composed of an amino acid part and a transfer ribosome nucleotide adenosine (tRNA). The aminoacylation adds an aminoacyl group to the compound, is as well related to the transport or transfer of an amino acid group to the fabric of proteins that is in fact the ribosome. In this manner, tRNA serves as a transportation means by adding the aminoacyl group to the tRNA molecule used in the biosynthesis of new protein (Cox and Nelson, 2013).

The current pathway involves different intermediates including glutamine, methionine, valine, alanine, lysine and the isoleucine. Increases in concentration of glutamine, methionine, valine, alanine, lysine, as detected by the heatmap from the list of the 26 top biomarkers, shows the importance of the aminoacyl-tRNA route and the subsequent production of proteins. Furthermore, the PCR technique could detect glutamine and isoleucine as major biomarkers in the NPC1 disease group compared to the wild treated cohorts. The aminoacyl-tRNA biosynthesis is linked to the protein production, with new protein biosynthesis taking place as a consequence of the tRNA adding an amino acid group to the new protein being constructed. The ^{1}H NMR metabolic profile of the NPC1 disease patients' blood plasma exhibited an increase of the 1-methylhistidine as confirmed by the PCR main biomarkers selection. This amino acid generation is usually linked to protein degradation, and as a consequence muscle mass loss, which is a sign of the NPC1 disease progression (Garver et al., 2007; Ruiz-Rodado et al., 2014). Therefore, the increase of glutamine levels for NPC1 disease subjects is consistent with protein degradation which, in turn, reduces the concentration of alanine in blood plasma. Indeed, increased levels of the concentration of amino acids is likely to be related to liver

parenchyma necrosis, which is connected to the hepatic fibrosis (Cox and Nelson, 2013; Garber, 1978; Probert et al., 2017).

Furthermore, BCAAs are important in terms of the regulation of the glucose levels in the bloodstream. After meals, the level of insulin and glucagon varies in the opposite direction, maintaining the sugar level balanced in blood. Hence, the elevated levels of BCAAs are in line with the muscle protein deterioration level (Garver et al., 2007; Ruiz-Rodado et al., 2014). Nevertheless, other pathways can assist in clarifying metabolic NPC1 disease changes.

- **Phenylalanine Metabolism**

Phenylalanine metabolism pathway is related to a phenylketonuria (PKU), an inborn error of metabolism, which is responsible for some form of neural development deficiency or mental retardation. In reality, phenylalanine can be converted to tyrosine by hydroxylation with the enzyme phenylalanine hydroxylase. The deficit of this intermediate can cause an excess of the phenylalanine. However, another pathway can be used for phenylalanine metabolism, and this involves reaction of phenylalanine with pyruvate to form the phenylpyruvate, which then accumulates in patients' blood (Bioinformatics, 2017). Under those circumstances, increases in phenylalanine and phenylpyruvate levels can exert grave consequences such as an advanced mental retardation due to slow brain development, which might be observed in some NPC1 patients developing the disease at an early age. Excess levels of amino acids, including phenylalanine can cause a deficiency in some metabolic routes. Consequently, phenylalanine and other amino acids being in excess in the patients' blood while competing to be shifted in the blood (Bioinformatics, 2017). In this respect, it has to be remembered that a study has shown that the level of phenylalanine in the blood is important as it is an indication of the abnormal development of the brain at younger age resulting in serious mental health problems for the NPC1 disease patients (Bioinformatics, 2017).

PCR has detected pyruvate as one of the main biomarkers while phenylalanine was seen to be important but to a lesser degree. This might be related to the fact that different model uses different algorithm that might better recognise a biomarker than another model (Les et al., 2016; McIntosh et al., 2016).

The next pathway which is very important in the understanding of the NPC1 disease aetiology is the glycolysis pathway that is next dissected.

- **Glycolysis and Gluconeogenesis**

Glycolysis as a metabolic pathway takes place in the cytosol of the molecular cells and does occur under aerobic or anaerobic (absence of oxygen) condition. This process of breaking down the glucose at different stages and the production of different intermediates such as the glucose-6-phosphate leads to the production of the pyruvate as end-product of the glycolysis. Lactate was detected by SVA and PCR models in higher concentration in the NPC1 disease patients' blood plasma over those of healthy controls individuals. The different ranking techniques featured it amongst the 3 major biomarkers detected. This attests to the importance of the glycolysis pathway in NPC1 disease aetiology. Indeed, the ***pyruvate*** can be converted back to glucose via the gluconeogenesis pathway, or using the ***acetyl-CoA*** route enabling the production of fatty acids. Therefore, it is understood that fat build-up at lysosomes and late endosome is related to the mutation of NPC1 genes, the potential lack of regulation of the trafficking of cholesterol, lipids such as sphingolipids due to defective NPC1 genes, ineffectiveness and inoperability can cause the aggravation of the NPC1 disease leading to cell death (Blom, 2003; Garver et al., 2007; Xu et al., 2010).

6.3. Discussion Based on NPC Liver Dysfunction Disease (NPC LDD) Dataset

6.3.1. Statistical Analysis and the NPC Disease Dataset

The severity of NPC1 disease could reach the extent where the patient's liver can be seriously affected. The following discussion is related to this next stage. The use of ITMTs to enhance our understanding of this disease development has been performed here through all the three models taken separately.

The OSVM technique could distinguish between the main classes of mice, i.e. those with the NPC-associated liver dysfunction disease (NPC LDD) and healthy control mice. The classification performances vary from one set of implementations to the next, in view of one model applying a specific algorithm, and hence achieving a particular performance. The setting using the RBF as a main kernel function and the AUC ROC as a classifier performance

measurement indicator was employed since the combination RBF-AUC ROC was found to perform more effectively than all other combinations, involving linear function and the Gmean performance metrics. Under those conditions, the average values obtained were AUC ROC = 0.973, std. deviation = 0.0048 and 95% confidence interval ±0.009449 (Table 30) are all consistent or improved on similar research performed (Martí and Reinelt, 2011; Tang et al., 2009). In a similar manner, the results also show that increasing the variable such as the number of time the stratified cross validation is repeated, the number of k folds used, and finally the application of standardisation, increases OSVM performance. The rationale behind this has been explained in subsection 6.2.1. These values attest the highest level achieved by the OSVM using a combination of the RBF-AUC ROC approach in discriminating between the two classes. The results are further corroborated by the scalar visualisation algorithm that is outlined below.

The SVA creates a relationship between scalar values or an index corresponding to the peak amplitude and the colour in the colour look-up table, and hence allows the matching between a scalar value and a colour in the look-up table. This, in turn, enables a visualisation of the changes in the concentration of the main metabolites involved in the NPC LDD. The change of colour, pattern, etc., can reveal a significance difference between a given metabolite and the rest of the metabolome involved in the study. In this manner, major or potential biomarkers can stand out from the different metabolites present in this dataset, and therefore aids our understanding of disease aetiology. In the NPC liver scale data plot (Figure 17) changes in intensity, colour and structure signify a major difference between the current feature and the remaining metabolites in the dataset. In the same manner, the contour plot (Figure 18) shows significant change in intensity and structure that might explain the NPC disease's main causes of progression. Thus, the column numbers 69, 68, 70, 67, 58, 59, 60, 55, 54, 17, 18, 7 (in decreasing order of importance) may indeed represent potential biomarkers for NPC liver dysfunction disease (NPC LDD) diagnosis. The corresponding chemical shifts are [3.81…3.87] [3.71…3.77] [3.39...3.44**]** [3.87…3.93] [3.77…3.81] [3.44…3.50] **[3.37…3.39]** [3.26…3.31] [3.21...3.26] [1.30…1.36] [1.45...1.51] [0.94...0.99] ppm. The change of colour is related to the importance in the disease aetiology, with green very important, light green fairly important and bleu less important. The respective metabolites detected are glycogen, the glutamate, glutamine, taurine, glycerophosphocholine, cysteine, myo-inositol, gamma-phosphorylcholine, lactate, alanine, leucine and isoleucine.

The markers discovered by the SVA strategy are therefore either very important biomarkers or potential ones for NPC LDD. For example, the glycogen that is the main form of storage of glucose in the liver of the human and animal cells (HMDB, 2017). In the isoleucine, valine and leucine biosynthesis, the three amino acids known as the branch-chain amino acids BCAAs. Their degradation involves the pyruvate that is in turn important in maintaining the glucose level in the blood. Additionally, the pyruvate is the anion and conjugate form of the pyruvic acid, with the latter producing through gluconeogenesis the fats. The fat storage for lack of gene providing the necessary transportation for degradation is one of the major source of the NPC LDD.

However, the PCR technique could detect the main biomarkers using the tri-ranking techniques developed. In the NPC LDD dataset, the following biomarkers were detected and ranked as very important in terms of the system studied. The 13 best biomarkers were methionine, hypotaurine, lysine, ornithine, phenylalanine, hippurate, nicotinate, xanthine, citrate, phosphorocholine, tyrosine, leucine, and isoleucine. It has to be remembered that pyruvate is present amongst the highly ranked (24 top) biomarkers detected by the PCR.

The model ranked methionine and hypotaurine as the best potential biomarkers in the NPC LDD dataset. Their actual relation to disease prognosis and diagnosis is therefore of paramount importance. Methionine metabolism has been shown to be related to two pathways, including the methionine cycle and the methionine trans-sulfuration. Different intermediates are necessary to produce homocysteine and adenosine. However, Trans sulfuration of homocysteine generates the cysteine that is converted into propionyl-CoA that is, in turn, transformed into succinyl-CoA (Cox and Nelson, 2013; SMPDB, 2015b). It should be noted that the cystine obtained from the oxidation of two molecules of cysteine is a more stable form of the latter. Current pathway controls are related to the availability of both methionine and cysteine, which in turn will favour production of cysteine and α-ketobutyrate, these being more glucogenic (Cox and Nelson, 2013; SMPDB, 2015b). In addition, the hypotaurine has also been shown to be a very important marker in the NPC-associated liver dysfunction (Hansen and Horslen, 2008; Mato et al., 2008; Ruiz Rodado et al., 2016).

Furthermore, ^{1}H NMR-detectable metabolites of mice liver dysfunction samples lysine and ornithine were classified 2nd best biomarkers. Lysine catabolism is achieved through a series of reactions connected to fatty acids degradation, and products of the lysine degradation

pathway produce ketones, especially under fasting conditions (Cox and Nelson, 2013; SMPDB, 2015c). Ornithine (classified amongst the 13 most significant biomarkers in the NPC LDD) can generate pyruvate or be transformed to gamma aminobutyrate. Pyruvate importance in the NPC1 disease dataset has been already established in this study (see 5.3.5). In a similar manner, its importance in the NPC-associated liver dysfunction has been noted both here and previously. For instance, the upregulated concentrations of lysine/ornithine in the NPC liver disease samples relative to those of wild treated (WT) was highlighted (Ruiz-Rodado et al., 2016).

Nevertheless, further statistics included in the present study, including a pathway's impact on an understanding of disease aetiology, the false detection rate (FDR), and the Holm adjusted p values provided further valuable information.

6.3.2. Pathway Analysis and NPC Liver Dysfunction

The pathway analysis for a given metabolite or group of metabolites is related to the biological pathway used to produce the molecules involved. Several molecular pathways analysis will be discussed, together with their involvement in the NPC liver dysfunction disease's diagnosis. Amongst these are alanine, aspartate and glutamate metabolism, phenylalanine metabolism, and arginine and proline metabolism, etc.

- **Alanine, Aspartate and Glutamate Metabolism**

The current pathway is highly differentially expressed, with elevated levels of the following metabolites, including N-acetyl-L-aspartate, aspartate that can be transformed to the fumarate through different intermediates. From alanine and aspartate, glutamate can be produced in combination with 2-oxoglutarate through the enzyme transaminase. The result of this reaction is the production of pyruvate and oxaloacetate, which are very important in glucose formation or degradation, i.e. gluconeogenesis or glycolysis (Cox and Nelson, 2013; SMPDB, 2015d). The importance of glucose in this pathway has been demonstrated, with the potential production of pyruvate and the reverse reaction producing glucose. The amination of pyruvate through alanine transaminase can produce the alanine. This reaction is one of the most used

means of generating alanine (Cox and Nelson, 2013; SMPDB, 2015e). However, aspartate synthesis in the human body is achieved by transamination of oxaloacetate using aspartate aminotransferase. The aspartate synthetised is used in protein production (Cox and Nelson, 2013; SMPDB, 2015f). Glutamine can be hydrolysed and transformed to glutamate (Glu) an alpha amino acid that can be synthetised by the human body. The latter can then be converted to gamma-aminobutyrate which, in turn, will produce the succinate. It should also be noted that glutamate plays a crucial role in the neural activation system in humans. This may explain certain conditions related to the NPC disease in general, and the stiffness of the NPC1 disease carrier.

- **Arginine and Proline Metabolism**

Arginine and proline metabolism give rise to production of arginine, ornithine, proline, citrulline and glutamate. In fact, the production of glutamate and proline from ornithine is an option available on the current pathway. For example, proline can be biosynthesised from glutamate and 1-pyrroline-5-carboxylate. Since the production of arginine, glutamine and proline can be bidirectional, depending on the cell type, the development processes, their stages, and concentrations vary significantly at the small intestinal level. Indeed, here, glutamine and glutamate produce citrulline that is extracted at the kidney level, i.e. removed from the circulation and transformed to arginine and then re-introduced to the circulation. Therefore, problem at kidney level will affect the arginine production, which in turn needs to be re-introduced in the in the circulation process. Arginine serves as precursor in the synthesis of various agents such as nitric oxide, creatine, polyamines, agmatine, and urea. Nitric oxide is believed to relax blood vessels (Frolkis et al., 2010; Jewison et al., 2014; Morris, 2007; SMPDB, 2010).

It should therefore be noted that proline and arginine are amino acids that are involved in protein synthesis using prolyl-tRNA and the arginyl-tRNA, themselves produced by their corresponding tRNA synthetases. However, at the liver level, ornithine is used in the cells as both an intermediate in the urea cycle, or for the synthesise of glutamate and glutamine, or glutamate and proline (Jewison et al., 2014; SMPDB, 2010). Indeed, ornithine is a major component involved in the current pathway, and is used in the liver, for the synthesis of glutamate and glutamine; however, in peripheral tissues it is involved in the synthesis of the glutamate and proline (Cox and Nelson, 2013).

- **Glutamine and Glutamate Metabolism**

The two main metabolites present in this pathway will be followed to understand the process of conversion of glutamine to glutamate. Glutamate is considered as a non-essential amino acid which can be produced by the human body. As a precursor molecule, it is involved in the production of different metabolites such as N-acetyl-L-glutamate, L-γ-glutamyl—L-cysteine, etc. It is also employed in the biosynthesis of certain amino acids, which include L-proline and L-arginine (Collard et al., 2010; Murphy et al., 1996; Yelamanchi et al., 2016).

Glutamate is produced from glutamine, alpha-ketoglutarate and 5-oxoproline. The synthesis of glutamate involves several reversible and irreversible reactions, which are catalysed by enzymes and the help of activators and inhibitors. For example, L-glutamine is transformed to N-acetyl-L-glutamate in the presence of N-acetylglutamate synthase, where L-arginine plays the role of activator and the inhibitors are succinate, coenzyme A, N-acetyl-L-aspartate and N-acetyl-L-glutamate (Shigesada and Tatibana, 1978; Yelamanchi et al., 2016).

However, the conversion of glutamine to glutamate requires enzymes such as glutaminase (GLS/GLS2), glutamine-fructose-6-phosphate transaminase (GFPT1 and GFPT2) and phosphoribosyl pyrophosphate amidotransferase (PPAT) (Hu et al., 2010; Yelamanchi et al., 2016). Glutamine is the precursor in reactions generating various metabolites such as N-acetyl-L-glutamate, L-γ-glutamyl-L-cysteine, δ-1-pyrroline-5-carboxylate, β-citrylglutamate (Battaglioli et al., 2003; Yelamanchi et al., 2016). Also, as a precursor of glutamate, glutamine plays a protective role in the human body essentially against nutrient depletion and certain tumour stresses. In a similar manner, enzymes related to glutamate metabolism or the alteration in glutamate metabolism are closely related to different other forms of neurodegenerative diseases (Burbaeva et al., 2005; Qi et al., 2013). The reversible transformation between glutamate and glutamine is presented through this research study as ones of the most significant pathways in NPC1 disease diagnosis. Therefore, alteration of glutamate metabolism will have significant effect on the NPC1 disease development.

This latter stage of the Discussion shows how the different pathways detected, and especially the biomarkers involved by the different models (ITMTs), are interconnected. In addition, they

highlight the importance of biomarkers discovered for NPC1 disease, and the more severe forms of the disease involving liver dysfunction.

6.4. Chapter Summary

Throughout the current discussion, the following results were attained and were related to the diagnosis of NPC1 disease, together with the more severe form of the disease affecting patients' liver (NPC LDD). This includes the points noted below;
1). Results from the optimum support vector machine (OSVM) classifier confirms that a combination of the RBF kernel and the AUC ROC approaches performed more effectively than the combinations involving the linear kernel and the Gmean techniques.

2). The OSVM model using a combination of RBF-AUC ROC could segregate the two classes of individuals in both the NPC1 disease and the NPC LDD datasets, achieving a very high level of performance, i.e. 0.944 and 0.973 respectively.

3). The scalar visualisation algorithm (SVA) could detect several biomarkers has been shown to be important in the NPC1 and the NPC LDD aetiologies. This included, hexacosanoate, (R)-3-hydroxybutyrate, L-fucose, the lactate, 3-hydroxyisovalerate, citrate, N-acetyl-4-O-acetylneuraminate, methionine, and glutamine for the NPC1 disease, and Glycogen, glutamate and glutamine, taurine, glycerophosphocholine, acetoacetate, taurine, myo-inositol, lactate, leucine, isoleucine, and alanine for NPC-associated liver dysfunction disease.

4). The PCR as last ITMTs model developed, and could as well detect several biomarkers that have been shown to be important in the NPC1 and the NPC LDD aetiologies. Other biomarkers also have been discovered the importance of which has not been emphasised in any previous research performed in relation to these conditions. This includes, the adipate and hippurate for NPC1 disease diagnosis, given that adipate/n-butyrate is a mixture of both molecules. Consequently, this biomarker can also be considered as important biomarkers in perhaps disease progression. In this NPC LDD study, most of the main biomarkers detected have already been discovered as important biomarkers in the NPC1/NPC LDD (both diseases being the same) disease study. This includes, methionine, hypotaurine, lysine, ornithine,

phenylalanine, nicotinate, xanthine, etc. Nevertheless, hippurate has not been formally categorised as probable main biomarker in the NPC liver associated with this disease. Finally, adipate and hippurate have been included as possible major biomarker for NPC1 disease diagnosis by the Computational Intelligence Techniques (CITs) based on the intelligent tri-modelling techniques (ITMTs) applied in this PhD thesis.

5). The intelligent tri-modelling techniques (ITMTs) combined with the tri-ranking techniques (TRTs) are valid strategy developed that have shown their importance, effectiveness and accuracies in the detection and ranking of biomarkers in the quest for a clear understanding of the NPC1 disease and its associated liver malfunction.

"The only thing you absolutely need to know, is the location of the library"

Albert Einstein

7. CONCLUSIONS - CONTRIBUTION TO KNOWLEDGE & FURTHER RESEARCH

7.1. Introduction

This chapter acts as a closing section to this PhD thesis, where a final summary of the research work's achievements are provided. Indeed, it highlights the contribution to knowledge; and also notes limitations of the research. This chapter also presents the further research planned in relation to the NPC1 disease research in general. Finally, an overall conclusion to the whole research investigations conducted is made.

7.2. Summary of Research

This research commenced by mining in a theoretical field of study to find a suitable one in order to give appropriate support to the present research. This allow to widen the researcher's own understanding of the research area. The notion of metabolomics was defined, together with related terms such as metabolites and the metabolomes. In addition, Computational Intelligence Techniques (CITs) based on the intelligent tri-modelling techniques (ITMTs) to be developed in this study were introduced, presenting some of the algorithms necessary to support this empirical research. The next chapter commenced by simply mentioning the notion of research methodology, while describing the terms 'approach' and 'theory'. Thus, the intelligent technology task fit model (ITTFM) was developed and supported the development process. Next, the research method employed was mentioned, with a combination of the

fieldwork and laboratory-based research being more adapted to this research. Although both datasets were supplied by collaborating Universities such as Oxford University (UK), etc., the method used was briefly presented. Furthermore, samples and feature selections were investigated given that some of the techniques were applied in the design process. The different challenges were identified, with methods employ to overcome the ones highlighted.

The next chapter related to the research design focussed on the mixed-experimental design, with data collection techniques and data analysis tools presented. To complete the design chapter, the technique to be applied in the validation stage was analysed, with notes on the dual validation processes addressed. This included a validation process related to the performance measurement based on the standard, a second validation procedure based on the theoretical framework developed in this thesis, a third validation is related to the data analysis technique utilised, and a final validation involving biomarkers discovered statistical validation. The result chapter and the related analysis encompasses two main sections, i.e. the results of the NPC1 disease and NPC liver dysfunction disease (NPC LDD) diagnosis, together with the different biomarkers and related pathways analysis. Furthermore, the classification algorithm OSVM could perfectly discriminate with higher performance using the RBF kernel the different groups that are NPC1 patients and wild treated/healthy control group with an AUC ROC value of 94.4%. However, the NPC liver dysfunction dataset was discriminated with an AUC ROC value of 97.3%. Chapter 6 involved a discussion section focussed on the metabolic featured pathways detected, and the main biomarkers placed on these pathways. How they are connected to one another in our understanding of the NPC1 disease's aetiology. In the final chapter which is related to the conclusion, the contribution to knowledge and further research, findings are dissected together with their importance in terms of NPC1 disease diagnosis. In this last chapter, recommendations are made regarding the possibility of carrying this research work further in the near future.

7.3. Contribution to Knowledge

Different contributions were made in different area of the PhD thesis. Indeed, contributions were made with relation to the research methodology applied including the research method applied with the development of the intelligent technology task fit model (ITTFM), and the

development of the tri-ranking techniques (TRTs). More importantly is the discovery of the biomarkers in the NPC1 and related diseases diagnosis, which could ultimately improve the research based on the NPC1 disease diagnosis in general. Henceforth,

- The development of the ITTFM allows us to streamline the different stages of the development process where different software and algorithms are involved. The sequences of research and selection of the Computational Intelligence Techniques (CITs) developed in this research supported by different algorithms such as the intelligence algorithm development for process validation purposes.

- The computational intelligence techniques (CITs) used in this research were based on the intelligent tri-modelling techniques (ITMTs) that are three different algorithms. Among the model included in the ITMTs are the optimum support vector machine (OSVM) for classification purpose and prediction, which is an optimal model of a conventional support vector machine. The scalar visualisation algorithm that, in turn, allows us to visualise the underlying transformation occurring at the molecular level, and which can explain the metabolic transformations. Finally, principal component regression (PCR) was presented as a combination of the principal component analysis (PCA) and multiple logistic regression (MLR). The objectives in using PCR in this research was to devise a strategy of detecting principal components in diseases analysis in view of strong correlations between the disease features in such a manner that the features can be correlated to the disease detection. Additionally, it allowed a ranking of disease features based on their importance in explaining the disease chemopathology.

- The tri-ranking techniques (TRTs) were developed to enable and improve the ranking of the diseases features based on their importance in its aetiology. In this manner, the top ranked features are considered to be the main biomarkers. The feature ranking techniques developed include the sum product of the coefficients (SPC), the exponential sum product of the coefficients (ESPC), and finally, the heuristic approach that gives a ranking favouring speed to the detriment of its precision and accuracy.

- The PCR model using the tri-ranking techniques have detected and ranked the following biomarkers as main biomarkers for NPC1 disease and the main complex form of the disease that is the NPC liver dysfunction associated with this disease. These are respectively the ***adipate*** and the ***hippurate.*** Furthermore, the intelligent tri-modelling techniques and the tri-ranking techniques have allowed to confirm previous biomarkers detected in other studies. This includes, pyruvate, glutamate, isoleucine,…etc., for NPC1 disease. Additionally, lactate, leucine, isoleucine, alanine, methionine, hypotaurine, lysine, ornithine, phenylalanine, nicotinate, and xanthine were confirmed as main biomarkers in NPC liver dysfunction disease (NPC LDD).

  Indeed, biomarkers such as methionine, hypotaurine, lysine, and phenylalanine have been detected as main biomarker in study conducted elsewhere (Mato et al., 2008; Ruiz-Rodado, 2016). In addition, validation procedures have been applied and explained in chapter 5, with regards to previous study conducted (Heron et al, 2013), which shows the validity of the techniques developed, and the research methodology applied.

7.4. Research Limitations

NPC1 disease lies amongst the restricted circle of rare diseases termed as 'orphan' diseases, where the investment for research is quite limited. This is ascribable to the fact that the number of known people with the condition is limited. Moreover, patients are still struggling in terms of being aware of their own conditions, making it difficult to know the exact number of patients and carry out upstream treatment in order to minimise the effect of the disease on them. The present study carries some limitations in view of several factors. Among these are:

- Possible biomarkers needing further investigation to confirm their status.
- Exponential sum product of the coefficients (ESPC) ranking technique can be further developed in terms of its mathematical formula.
- Involvement of negative coefficients, including the negative coefficients of regression and correlation may be important in features ranking. The present techniques did not emphasise on this aspect thoroughly. Indeed, it can be demonstrated that the greater the absolute value of the coefficients of regression and correlation the more important the

effect on the ranking result. This can be performed by introducing the technique known as the Absolute Value of the Sum Product of the Coefficients (AVoSPC), which was not applied in this research (included in further research).

- The biomarkers detected by the scalar visualisation does not in all the cases correspond to an exact number, hence the biomarker detected might not be precise.

7.5. Further Research Proposed

In the light of the present study, the possibility of taking it forward for future research is available. Some of these possibilities are outlined below with their main and planned direction to follow highlighted.

- Further mathematical development of the exponential sum product of the coefficients (ESPC), in which the terms related to feature contributions as exponential terms can be further developed.
- The development of the Absolute Value of the Sum Product of the coefficient (AVoSPC) techniques. This further development may favour more the negative terms, and hence should be regulated with appropriate coefficient or weightings ascribed to feature contributions.
- Application of these techniques and models developed in the drug development field, especially those designed for NPC1 disease treatment and secondary conditions.
- The development of sound biomarkers validation technique for the ones discovered, including pyruvate and adipate.

7.6. Conclusion

This research was focused on the detection of biomarkers in order to improve our understanding of the chemopathogenesis of NPC1 disease. Major biomarkers were found

applying different intelligent modelling techniques supported by Computational Intelligence Techniques (CITs). Biomarkers discovered included well-known ones and the others not strongly connected to present disease development. In addition, ranking techniques were developed. Amongst the biomarkers discovered some well-known ones, and also new ones were detected, including adipate and hippurate. Biomarkers detection techniques developed were validated by standard performance analysis methods such as combining kernel RBF function and ROC-AUC as performance measure. However, different techniques generated different ranking for biomolecules detected. In this regard, biomarker discovery through sphingolipids profiling has provided evidence that the pathogenesis of the NPC1 disease is very complex and therefore the use of a given biomarker or class of biomarkers is most unlikely to provide a clear solution. Therefore, tackling this disease using different approaches i.e., targeting different aspects of the pathology, might be more productive in terms of effectiveness in its diagnosis (Fan et al., 2013a). Such biomarkers discovery may play a key role again in the design of therapeutic agents for its treatment.

REFERENCES

Abbott, R.D., Carroll, R.J., (1984). Interpreting multiple logistic regression coefficients in prospective observational studies. Am. J. Epidemiol. 119, 830–836.

Abraham, G., Havulinna, A.S., Bhalala, O.G., Byars, S.G., De Livera, A.M., Yetukuri, L., Tikkanen, E., Perola, M., Schunkert, H., Sijbrands, E.J., Palotie, A., Samani, N.J., Salomaa, V., Ripatti, S., Inouye, M., (2016). Genomic prediction of coronary heart disease. Eur. Heart J. 37, 3267–3278. doi:10.1093/eurheartj/ehw450

Adetokunbo, O.L., Roy, M.A., Barry, R.B., Sissela, B., Lincoln, C.C., (2011). Bulletin of the World Health Organization.

Akay, M.F., (2009). Support vector machines combined with feature selection for breast cancer diagnosis. Expert Syst. Appl. 36, 3240–3247. doi:10.1016/j.eswa.(2008).01.009

Akst, J., (2014). Rare Disease ti Inform Ebola Susceptibility. The Scientist.

Albro, M.B., Li, R., Banerjee, R.E., Hung, C.T., Ateshian, G.A., (2010). Validation of theoretical framework explaining active solute uptake in dynamically loaded porous media. J. Biomech. 43, 2267–2273. doi:10.1016/j.jbiomech.(2010).04.041

AMATH 301, (2016). Lecture: PCA for Face Recognition.

Amathieu, R., (2016). Nuclear magnetic resonance based metabolomics and liver diseases: Recent advances and future clinical applications. World J. Gastroenterol. 22, 417. doi:10.3748/wjg.v22.i1.417

Anton, H., (1987). Elementary linear algebra. Wiley, New York.

Artemiou, A., Li, B., (2009). On Principal Components and Regression: A Statistical Explanation of a Natural Phenomenon. Stat. Sin. 1557–1565.

Balch, W.E., Holson, E., Ory, D.S., Parmacek, M.S., Patterson, M.C., Pavan, W.J., Pfeffer, S., (2008). About Niemann-Pick Type C Cause, Diagnosis, Symptoms and Treatment.

Balunas, M.J., Kinghorn, A.D., (2005). Drug discovery from medicinal plants. Life Sci. 78, 431–441. doi:10.1016/j.lfs.(2005).09.012

Bapir, M.A., (2010). Is it possible for qualitative research to be properly valid and reliable. Univ. Warwick.

Bartel, J., Krumsiek, J., Theis, F.J., (2013). Statistical Methods for the Analysis of High-throughput Metabolomics Data. Comput. Struct. Biotechnol. J. 4, 1–9. doi:10.5936/csbj.201301009

Becker, S., Kortz, L., Helmschrodt, C., Thiery, J., Ceglarek, U., (2012). LC–MS-based metabolomics in the clinical laboratory. J. Chromatogr. B 883–884, 68–75. doi:10.1016/j.jchromb.(2011).10.018

Beecher, C.W., (2003). The human metabolome, in: Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis. Springer, pp. 311–319.

Beger, R., (2013). A Review of Applications of Metabolomics in Cancer. Metabolites 3, 552–574. doi:10.3390/metabo3030552

Beger, R.D., Sun, J., Schnackenberg, L.K., (2010). Metabolomics approaches for discovering biomarkers of drug-induced hepatotoxicity and nephrotoxicity. Toxicol. Appl. Pharmacol. 243, 154–166. doi:10.1016/j.taap.(2009).11.019

Bell, J., (2014). Doing Your Research Project: A Guide for First-Time Researchers. McGraw-Hill Education (UK).

Bi, X., Liao, G., (2010). Cholesterol in Niemann–Pick Type C disease, in: Harris, J.R. (Ed.), Cholesterol Binding and Cholesterol Transport Proteins: Springer Netherlands, Dordrecht, pp. 319–335.

Bioinformatics, (2017). Chapter 17 : Amino Acid Oxidation and the Production of Urea [WWW Document]. Bioinfo.org.cn. URL http://www.bioinfo.org.cn/book/biochemistry/chapt17/sim5.htm (accessed 6.3.17).

Bird, V., Leamy, M., Tew, J., Le Boutillier, C., Williams, J., Slade, M., (2014). Fit for purpose? Validation of a conceptual framework for personal recovery with current mental health consumers. Aust. N. Z. J. Psychiatry 0004867413520046.

Blagus, R., Lusa, L., (2012). Evaluation of SMOTE for High-Dimensional Class-Imbalanced Microarray Data. IEEE, pp. 89–94. doi:10.1109/ICMLA.(2012).183

Blom, T.S., (2003). Defective endocytic trafficking of NPC1 and NPC2 underlying infantile Niemann-Pick type C disease. Hum. Mol. Genet. 12, 257–272. doi:10.1093/hmg/ddg025

Boardman, M., Trappenberg, T., (2006). A heuristic for free parameter optimization with support vector machines, in: Neural Networks, (2006). IJCNN'06. International Joint Conference On. IEEE, pp. 610–617.

Boston University School of Public Health, (2013). Multiple Logistic Regression Analysis [WWW Document]. URL http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_multivariable/BS704_Multivariable8.html (accessed 4.17.17).

Brereton, R.G., (2012). Self-organising maps for visualising - modelling. Chem. Cent. J. 6, 1.

Bruker Corporation, (2013). Bruker – Innovation with Integrity.

Bruker Corporation [WWW Document], n.d. www.bruker.com. URL https://www.bruker.com/nc/search.html?q=NMR+spectrometer+features&tx_ptsolr_pi1%5Bsolr%5D%5Bfilters%5D%5BfacetFilterbox%5D%5BbereichFacetFilter%5D%5BfilterValues%5D=&resetPager=1 (accessed 5.4.15).

Buduma, N., (2015). Data Science 101: Preventing Overfitting in Neural Networks [WWW Document]. URL http://www.kdnuggets.com/(2015)/04/preventing-overfitting-neural-networks.html (accessed 6.10.17).

Burns, R.B., (2000). Introduction to research methods, 4th ed. ed. SAGE, London ; Thousand Oaks, Calif.

Cateni, S., Vannucci, M., Vannocci, M., Coll, V., (2013). Variable Selection and Feature Extraction Through Artificial Intelligence Techniques, in: Freitas, L. (Ed.), Multivariate Analysis in Management, Engineering and the Sciences. InTech.

Chau, C.H., Rixe, O., McLeod, H., Figg, W.D., (2008). Validation of Analytic Methods for Biomarkers Used in Drug Development. Clin. Cancer Res. 14, 5967–5976. doi:10.1158/1078-0432.CCR-07-4535

Clough, P., Nutbrown, C., (2012). A Student's Guide to Methodology. SAGE.

Cluzeau, C.V.M., Watkins-Chow, D.E., Fu, R., Borate, B., Yanjanin, N., Dail, M.K., Davidson, C.D., Walkley, S.U., Ory, D.S., Wassif, C.A., Pavan, W.J., Porter, F.D., (2012). Microarray expression analysis and identification of serum biomarkers for Niemann-Pick disease, type C1. Hum. Mol. Genet. 21, 3632–3646. doi:10.1093/hmg/dds193

Cook, R.D., (2007). Fisher Lecture: Dimension Reduction in Regression. Stat. Sci. 22, 1–26. doi:10.1214/088342306000000682

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Cox, M., Nelson, D.L., (2013). Lehninger Principles of Biochemistry, 6th Revised edition edition. ed. Palgrave Macmillan, New York.

Creswell, J.W., (2003). Research design: qualitative, quantitative, and mixed method approaches, 2nd ed. ed. Sage Publications, Thousand Oaks, Calif.

Cuadros, D.F., Branscum, A.J., Crowley, P.H., (2011). HIV-malaria co-infection: effects of malaria on the prevalence of HIV in East sub-Saharan Africa. Int. J. Epidemiol. 40, 931–939. doi:10.1093/ije/dyq256

Daniel, S.O., Forbes, D.P., (2015). Methods of Determining Efficacy of Therapy For Niemann-Pick C Disease And Related Disorders US 20150226757 A1. US 20150226757 A1.

David, C., Matthieu, B., Lars, B., Mathieu, B., Noel, D., Kyle, K., Manoj, K., (2016). RBF SVM parameters [WWW Document]. URL http://scikit-learn.org/stable/auto_examples/svm/plot_RBF_parameters.html (accessed 11.5.16).

de Graaf, R.A., (2007). In vivo NMR spectroscopy: principles and techniques, 2nd ed. ed. John Wiley & Sons, Chichester, West Sussex, England ; Hoboken, NJ.

DeLisle, R., (2007). SOM tutorial part one [WWW Document]. Ai-Junkie. URL http://www.ai-junkie.com/ann/som/som5.html (accessed 12.11.16).

Dettmer, K., Aronov, P.A., Hammock, B.D., (2007). Mass spectrometry-based metabolomics. Mass Spectrom. Rev. 26, 51–78. doi:10.1002/mas.20108

Duarte, I.F., Diaz, S.O., Gil, A.M., (2014). NMR metabolomics of human blood and urine in disease research. J. Pharm. Biomed. Anal. 93, 17–26. doi:10.1016/j.jpba.(2013).09.025

Dudovsky, J., (2011). Research Approach - Research-Methodology [WWW Document]. URL http://research-methodology.net/research-methodology/research-approach/ (accessed 9.1.16).

Duke, (2014). Research Question Original.

European Commission, (2010). Biomarkers for Patients Stratification. European Commission.

Fan, M., Sidhu, R., Fujiwara, H., Tortelli, B., Zhang, J., Davidson, C., Walkley, S.U., Bagel, J.H., Vite, C., Yanjanin, N.M., Porter, F.D., Schaffer, J.E., Ory, D.S., 2013a. Identification of Niemann-Pick C1 disease biomarkers through sphingolipid profiling. J. Lipid Res. 54, 2800–2814. doi:10.1194/jlr.M040618

Fidock, D.A., Rosenthal, P.J., Croft, S.L., Brun, R., Nwaka, S., (2004). Antimalarial drug discovery: efficacy models for compound screening. Nat. Rev. Drug Discov. 3, 509–520. doi:10.1038/nrd1416

Fiehn, O., (2006). Metabolite profiling in Arabidopsis. Arab. Protoc. 439–447.

Fiehn, O., (2002). Metabolomics–the link between genotypes and phenotypes. Plant Mol. Biol. 48, 155–171.

Freundlich, M., Burns, R.O., Umbarger, H.E., 1962. Control of isoleucine, valine, and leucine biosynthesis, I. Multi-valent repression. Proc. Natl. Acad. Sci. 48, 1804–1808.

Frolkis, A., Knox, C., Lim, E., Jewison, T., Law, V., Hau, D.D., Liu, P., Gautam, B., Ly, S., Guo, A.C., Xia, J., Liang, Y., Shrivastava, S., Wishart, D.S., (2010). SMPDB: The Small Molecule Pathway Database. Nucleic Acids Res. 38, D480–D487. doi:10.1093/nar/gkp1002

Fu, G.-H., Zhang, B.-Y., Kou, H.-D., Yi, L.-Z., (2016). Stable biomarker screening and classification by subsampling-based sparse regularization coupled with support vector machines in metabolomics. Chemom. Intell. Lab. Syst. doi:10.1016/j.chemolab.(2016).11.006

Gabriel, D., (2011). Methods and methodology | Dr Deborah Gabriel.

Garver, W.S., Francis, G.A., Jelinek, D., Shepherd, G., Flynn, J., Castro, G., Walsh Vockley, C., Coppock, D.L., Pettit, K.M., Heidenreich, R.A., Meaney, F.J., (2007). The National Niemann–Pick C1 disease database: Report of clinical features and health problems. Am. J. Med. Genet. A. 143A, 1204–1211. doi:10.1002/ajmg.a.31735

Gauderman, W.J., Murcray, C., Gilliland, F., Conti, D.V., (2007). Testing association between disease and multiple SNPs in a candidate gene. Genet. Epidemiol. 31, 383–395. doi:10.1002/gepi.20219

Godoy, M.M.G., Lopes, E.P.A., Silva, R.O., Hallwass, F., Koury, L.C.A., Moura, I.M., Gonçalves, S.M.C., Simas, A.M., (2010). Hepatitis C virus infection diagnosis using metabonomics: Hepatitis C diagnosis using metabonomics. J. Viral Hepat. 17, 854–858. doi:10.1111/j.1365-2893.(2009).01252.x

Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G., Kell, D.B., (2004). Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol. 22, 245–252. doi:10.1016/j.tibtech.(2004).03.007

Grand Canyon University, (2015). Sampling Methods.

Grootveld, M., (2014). Metabolic Profiling: Disease and Xenobiotics. Royal Society of Chemistry.

Guo, Z., Sheffield, J., (2006). A paradigmatic and methodological examination of KM research: (2000) to (2004), in: System Sciences, (2006). HICSS'06. Proceedings of the 39th Annual Hawaii International Conference On. IEEE, p. 153a–153a.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., (2002). Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–422.

Hackstadt, A.J., Hess, A.M., (2009). Filtering for increased power for microarray data analysis. BMC Bioinformatics 10, 11. doi:10.1186/1471-2105-10-11

Halouska, S., Powers, R., (2006). Negative impact of noise on the principal component analysis of NMR data. J. Magn. Reson. 178, 88–95. doi:10.1016/j.jmr.(2005).08.016

Hansen, K., Horslen, S., (2008). Metabolic liver disease in children. Liver Transpl. 14, 713–733. doi:10.1002/lt.21520

Harrigan, G.G., Goodacre, R., (2003). Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function ... - Google Books, illustrated. ed. Springer Science & Business Media.

Hastie, T., Tibshirani, R., Friedman, J., (2008). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. ed. Elsevier.

He, H., Bai, Y., Garcia, E., Li, S., others, (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning, in: Neural Networks, (2008). IJCNN (2008). IEEE International Joint Conference On. IEEE, p. 1322–1328.

He, Q., Kong, X., Wu, G., Ren, P., Tang, H., Hao, F., Huang, R., Li, T., Tan, B., Li, P., Tang, Z., Yin, Y., Wu, Y., (2009). Metabolomic analysis of the response of growing pigs to dietary l-arginine supplementation. Amino Acids 37, 199–208. doi:10.1007/s00726-008-0192-9

Heppner, P.P., Kivlighan, D.M., Wampold, B.E., (2007). Research design in counselling, 3rd ed.

Heron, B., V., Valayannopoulos, V., Baruteau, J., Chabrol, B., Ogier, H., Latour, P., Dobbelaere, D., Eyer, D., Labarthe, F., Maurey, H., (2012). Miglustat therapy in the French cohort of paediatric patients with Niemann-Pick disease type C. Orphanet journal of rare disease, Volume 7 Issue 1.

Higgins, J., (2005). Introduction to Multiple Regression - bcg_comp_chapter4.pdf.

HMDB, (2015). Human Metabolome Database: Showing metabocard for Glycogen (HMDB00757) [WWW Document]. URL http://www.hmdb.ca/metabolites/HMDB00757 (accessed 6.10.17).

Hou, Y., Hossain, G.S., Li, J., Shin, H., Liu, L., Du, G., Chen, J., (2016). Two-Step Production of Phenylpyruvic Acid from L-Phenylalanine by Growing and Resting Cells of Engineered Escherichia coli: Process Optimization and Kinetics Modelling. PLOS ONE 11, e0166457. doi:10.1371/journal.pone.0166457

Hughes, C., (2012). Qualitative and quantitative approaches.

Hughes-Wilson, W., Palma, A., Schuurman, A., Simoens, S., (2012). Paying for the Orphan Drug System: break or bend? Is it time for a new evaluation system for payers in Europe to take account of new rare disease treatments? Orphanet J. Rare Dis. 7, 74. doi:10.1186/1750-1172-7-74

Jajuga, K., Walesiak, M., (2000). Standardisation of Data Set under Different Measurement Scales, in: Decker, P.D.R., Gaul, P.D.W. (Eds.), Classification and Information Processing at the Turn of the Millennium, Studies in Classification, Data Analysis, and Knowledge Organization. Springer Berlin Heidelberg, pp. 105–112. doi:10.1007/978-3-642-57280-7_11

Jewison, T., Su, Y., Disfany, F.M., Liang, Y., Knox, C., Maciejewski, A., Poelzer, J., Huynh, J., Zhou, Y., Arndt, D., Djoumbou, Y., Liu, Y., Deng, L., Guo, A.C., Han, B., Pon, A., Wilson, M., Rafatnia, S., Liu, P., Wishart, D.S., (2014). SMPDB 2.0: Big Improvements to the Small Molecule Pathway Database. Nucleic Acids Res. 42, D478–D484. doi:10.1093/nar/gkt1067

Johnson, C.R., (2015). Visualization, in: Engquist, B. (Ed.), Encyclopedia of Applied and Computational Mathematics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1537–1546. doi:10.1007/978-3-540-70529-1_368

Johnsona, C.R., (2012). Visualization of Scalar and Vector Fields.

Jonker, J., Pennink, B., (2009). The Essence of Research Methodology. Springer Berlin Heidelberg, Berlin, Heidelberg.

Joshi, D.D., (2012). Herbal Drugs and Fingerprints: Evidence Based Herbal Drugs. Springer Science & Business Media.

Keeler, J., (2004). Fourier Transformation and Data Processing [WWW Document].

Kim, J.-H., (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. Comput. Stat. Data Anal. 53, 3735–3745. doi:10.1016/j.csda.(2009).04.009

King, M.W., (2017). Nitrogen Metabolism and the Urea Cycle [WWW Document]. URL https://themedicalbiochemistrypage.org/nitrogen-metabolism.php (accessed 5.28.17).

Kohavi, R., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Int. Join Conf. Artif. Intell.

Köknar-Tezel, S., Latecki, L.J., (2011). Improving SVM classification on imbalanced time series data sets with ghost points. Knowl. Inf. Syst. 28, 1–23. doi:10.1007/s10115-010-0310-3

Komura, T., (2016). Scalar Algorithms: Colour Mapping.

Kosmides, A.K., Kamisoglu, kubra, Calvano, S.E., Corbett, S.A., Androulakis, I.P., (2013). Metabolomic Fingerprinting: Challenges and Opportunities. NIH 1–22.

Kotsiantis, S., Kanellopoulos, D., Pintelas, P., others, (2006). Handling imbalanced datasets: A review. GESTS Int. Trans. Comput. Sci. Eng. 30, 25–36.

Kruger, N.J., Troncoso-Ponce, M.A., Ratcliffe, R.G., (2008). 1H NMR metabolite fingerprinting and metabolomic analysis of perchloric acid extracts from plant tissues. Nat. Protoc. 3, 1001–1012. doi:10.1038/nprot.(2008).64

Kuhn, M., Johnson, K., (2013). Applied Predictive Modelling. Springer.

Kumar, G., Kalra, R., (2016). Decision Support System for Diagnosis of Heart Disease using PCA and SVM Classifier. Int. J. Recent Res. Asp. 3, 181–187.

Kumar, K.S., Madhu, G., (2012). Analysis and multinomial logistic regression modelling of work stress in manufacturing industries in Kerala, India. Int. J. Adv. Eng. Technol. 2, 410.

Kumar, R., (2011). A Step-by step Guide for Beginners, 3 rd Edition. ed. SAGE.

Kumar, S., nd. Selecting a Sample.

Larose, D.T., (2006). Data mining methods and models. Wiley-Interscience, Hoboken, NJ.

LASER, (2000). Chapter 3 - Sampling.

Lau, S., Lee, K.-C., Lo, G., Ding, V., Chow, W.-N., Ke, T., Curreem, S., To, K., Ho, D., Sridhar, S., Wong, S., Chan, J., Hung, I., Sze, K.-H., Lam, C.-W., Yuen, K.-Y., Woo, P., (2016). Metabolomic Profiling of Plasma from Melioidosis Patients Using UHPLC-QTOF MS Reveals Novel Biomarkers for Diagnosis. Int. J. Mol. Sci. 17, 307. doi:10.3390/ijms17030307

LeCompte, M.D., Goets, J.P., 1982. Problem of Reliability and Validity in Ethnographic Research', Review of Education Research, (Review of Educational Research).

Li, C., Wang, X., Dong, W., Yan, J., Liu, Q., Zha, H., (2015). Active Sample Learning and Feature Selection: A Unified Approach. ArXiv Prepr. ArXiv150301239.

Liang, X.-T., Fang, W.-S., (2006). Medicinal Chemistry of Bioactive Natural Products. John Wiley & Sons.

Lin, W.-J., Chen, J.J., (2013). Class-imbalanced classifiers for high-dimensional data. Brief. Bioinform. 14, 13–26. doi:10.1093/bib/bbs006

Mahmoodabadi, S.Z., Alirezaie, J., Babyn, P., Kassner, A., Widjaja, E., (2008). PCA-SGA implementation in classification and disease specific feature extraction of the brain MRS signals, in: (2008) 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 3526–3529.

Maji, S., Berg, A.C., Malik, J., (2013). Efficient Classification for Additive Kernel SVMs. IEEE Trans. Pattern Anal. Mach. Intell. 35, 66–77. doi:10.1109/TPAMI.(2012).62

Marcello Manfredi, E.R., (2013). Biomarkers Discovery through Multivariate Statistical Methods: A Review of Recently Developed Methods and Applications in Proteomics. J. Proteomics Bioinform. s3. doi:10.4172/jpb.S3-003

Martí, R., Reinelt, G., (2011). The Linear Ordering Problem, Applied Mathematical Sciences. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-642-16729-4

Massy, W.F., 1965. Principal Components Regression in Exploratory Statistics Research. J. Am. Stat. Assoc. 60, 234–256.

Mathieson, T., (2013). Niemann-Pick Disease Group UK.

MathWorks, 2016a. Trainbu: Batch Unsupervised Weight/Bias Training.

MathWorks, 2016b. MATLAB - The Language of Technical Computing [WWW Document]. URL

https://uk.mathworks.com/help/matlab/index.html;jsessionid=6b5f68e6274435cb641220ee2a90?/access/helpdesk/help/techdoc/learn_matlab/f3-25097.html= (accessed 2.28.17).

Mato, J.M., Martinez-Chantar, M.L., Lu, S.C., (2008). Methionine Metabolism and Liver Disease. Annu. Rev. Nutr. 28, 273–293. doi:10.1146/annurev.nutr.28.061807.155438

McCullagh, P., Yang, J., (2006). Stochastic classification models, in: International Congress of Mathematicians. p. 72.

McDermott, J.E., Wang, J., Mitchell, H., Webb-Robertson, B.-J., Hafen, R., Ramey, J., Rodland, K.D., (2013). Challenges in biomarker discovery: combining expert insights with statistical analysis of complex omics data. Expert Opin. Med. Diagn. 7, 37–51. doi:10.1517/17530059.(2012).718329

McDonald, J.H., (2015). Simple logistic regression - Handbook of Biological Statistics [WWW Document]. Handb. Biol. Stat. URL http://www.biostathandbook.com/simplelogistic.html (accessed 7.21.16).

McGregor, S.L., Murnane, J.A., (2010). Paradigm, methodology and method: Intellectual integrity in consumer scholarship. Int. J. Consum. Stud. 34, 419–427.

MEDICI, E., (2004). How to investigate the use of medicines by consumers 98.

MedicineNet, (2016). Biofluids Definition. Med. Dict.

Meekings, K.N., Williams, C.S.M., Arrowsmith, J.E., (2012). Orphan drug development: an economically viable strategy for biopharma R& D. Drug Discov. Today 17, 660–664. doi:10.1016/j.drudis.(2012).02.005

Mountrakis, G., Im, J., Ogole, C., (2011). Support vector machines in remote sensing: A review. ISPRS J. Photogramm. Remote Sens. 66, 247–259. doi:10.1016/j.isprsjprs.(2010).11.001

National Institute of Neurological Disorder and Stroke, (2016). Niemann-Pick Disease Information Page: National Institute of Neurological Disorders and Stroke (NINDS) [WWW Document]. NIH. URL http://www.ninds.nih.gov/disorders/niemann/niemann.htm (accessed 7.20.16).

Nayyar, G.M., Breman, J.G., Newton, P.N., Herrington, J., (2012). Poor-quality antimalarial drugs in Southeast Asia and sub-Saharan Africa. Lancet Infect. Dis. 12, 488–496. doi:10.1016/S1473-3099(12)70064-6

Ngo, L.T., Okogun, J.I., Folk, W.R., (2013). 21st Century natural product research and drug development and traditional medicines. Nat. Prod. Rep. 30, 584. doi:10.1039/c3np20120a

Nguyen, M.H., de la Torre, F., (2010). Optimal feature selection for support vector machines. Pattern Recognit. 43, 584–591. doi:10.1016/j.patcog.(2009).09.003

Niemann-Pick Disease Overview [WWW Document], (2013). . Natl. Niemann-Pick Dis. Found. Inc. URL http://www.nnpdf.org/npdisease_01.html (accessed 7.20.16).

Niemann-Pick UK, (2012). Treatment and Therapies - Type C - Niemann-PickUK. Niemann-Pick UK.

Nikas, J.B., Low, W.C., (2011). ROC-supervised principal component analysis in connection with the diagnosis of diseases. Am. J. Transl. Res. 3, 180–96.

Nilsen, M.M., Meier, S., Andersen, O.K., Hjelle, A., (2011). SELDI–TOF MS analysis of alkylphenol exposed Atlantic cod with phenotypic variation in gonadosomatic index. Mar. Pollut. Bull. 62, 2507–2511. doi:10.1016/j.marpolbul.(2011).08.006

Nwankwo, V., (2004). Research design and methodology. Fourth Dimension.

O'Leary, Z., (2013). The Essential Guide to Doing Your Research Project. SAGE.

Oxford Dictionaries, (2016). Definition of Saliva. Oxf. Dictionaries - Lang. Matters.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., others, (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peltonen, J., (2014). Dimensionality Reduction and Visualization.

Peng, Z., (2011). Interactive Visualization of Computational Fluid Dynamics Data.

Pineda M, Wraith JE, Mengel E, Sedel F, Hwu WL, Rohrbach M, Bembi B,Walterfang M, Korenke GC, Marquardt T, et al: Miglustat in patients with Niemann-Pick disease Type C (NP-C): a multicentre observational retrospective cohort study. Mol Genet Metab (2009), 98: 243–249.

Pratiwi, D., (2012). The use of self-organizing map method and feature selection in image database classification system. ArXiv Prepr. ArXiv12060104.

Pratiwi, T.A., (2013). Data Collection Techniques [WWW Document]. URL http://s3.amazonaws.com/academia.edu.documents/36947831/INTERVIEW_JADI.docx?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1478964599&Signature=OTQGHTMEt6Dh%2FfkvQ91gZgb3%2Bic%3D&response-content-disposition=attachment%3B%20filename%3DDATA_COLLECTION_TECHNIQUE_INTERVIEWING.docx (accessed 11.12.16).

PRIYAN, S., (2012). Methods of data collection [WWW Document]. URL http://www.slideshare.net/priyansakthi/methods-of-data-collection-16037781?next_slideshow=1

Pubchem, (2017). Pyruvic acid | C3H4O3 - PubChem [WWW Document]. URL https://pubchem.ncbi.nlm.nih.gov/compound/Pyruvic_acid (accessed 6.9.17).

Qi, S., Ouyang, X., Wang, L., Peng, W., Wen, J., Dai, Y., (2012). A Pilot Metabolic Profiling Study in Serum of Patients with Chronic Kidney Disease Based on ^{1} H-NMR-Spectroscopy. Clin. Transl. Sci. 5, 379–385. doi:10.1111/j.1752-8062.(2012).00437.x

Qi, S.-W., Tu, Z.-G., Peng, W.-J., Wang, L.-X., Ou-Yang, X., Cai, A.-J., Dai, Y., (2012). H NMR-based serum metabolic profiling in compensated and decompensated cirrhosis. World J Gastroenterol 18, 285–290.

Qiu, Y., Rajagopalan, D., Connor, S.C., Damian, D., Zhu, L., Handzel, A., Hu, G., Amanullah, A., Bao, S., Woody, N., MacLean, D., Lee, K., Vanderwall, D., Ryan, T., (2008). Multivariate classification analysis of metabolomic data for candidate biomarker discovery in type 2 diabetes mellitus. Metabolomics 4, 337–346. doi:10.1007/s11306-008-0123-5

Raschkas, S., (2014). About Feature Scaling and Normalization. Website [WWW Document]. URL sebastianraschka.com/Articles/(2014)_about_feature_scaling.html (accessed 2.5.17).

Ravanbakhsh, S., Liu, P., Bjorndahl, T.C., Mandal, R., Grant, J.R., Wilson, M., Eisner, R., Sinelnikov, I., Hu, X., Luchinat, C., Greiner, R., Wishart, D.S., (2015). Accurate Fully-Automated NMR Spectral Profiling for Metabolomics. PLOS ONE. doi:10.1371/journal.pone.0132873

Ray, S., (2015). Understanding Support Vector Machine algorithm from examples (along with code). Anal. Vidhya.

Research Rundowns, (2009). Instrument, Validity, Reliability. Res. Rundowns.

Rocha, C.M., Carrola, J., Barros, A.S., Gil, A.M., Goodfellow, B.J., Carreira, I.M., Bernardo, J., Gomes, A., Sousa, V., Carvalho, L., Duarte, I.F., (2011). Metabolic Signatures of Lung Cancer in Biofluids: NMR-Based Metabonomics of Blood Plasma. J. Proteome Res. 10, 4314–4324. doi:10.1021/pr200550p

Ronneberg, (2015). Regression - High p value Based on Residual Deviance when Model Appears to have Poor Fit - Cross Validated [WWW Document]. Stack Exchange. URL https://stats.stackexchange.com/questions/136021/high-p-value-based-on-residual-deviance-when-model-appears-to-have-poor-fit (accessed 5.17.17).

Rosario, S.F., Thangadurai, K., (2015). RELIEF: Feature Selection Approach. Int. J. Innov. Res. Dev. ISSN 2278–0211 4.

Ross, S.M., Morrison, G.R., 1996. Experimental research methods. Handb. Res. Educ. Commun. Technol. Proj. Assoc. Educ. Commun. Technol. 1148–1170.

Ruiz-Rodado, V., (2016). New Developments in 1H NMR-linked Metabolomics: Identification of New Biomarkers for the Metabolomic Classification of Niemann-Pick Disease, Type C1, and its Response to Treatment.

Ruiz-Rodado, V., Marcos Luque-Baena, R., te Vruchte, D., Probert, F., H. Lachmann, R., J. Hendriksz, C., E. Iraith, J., Imrie, J., Elizondo, D., Sillence, D., Clayton, P., M. Platt, F., Grootveld, M., (2014). 1H NMR-Linked Urinary Metabolic Profiling of Niemann-Pick Class

C1 (NPC1) Disease: Identification of Potential New Biomarkers using Correlated Component Regression (CCR) and Genetic Algorithm (GA) Analysis Strategies. Curr. Metabolomics 2, 88–121.

Ruiz-Rodado, V., Nicoli, E.-R., Probert, F., Smith, D.A., Morris, L., Wassif, C.A., Platt, F.M., Grootveld, M., (2016). 1H NMR-Linked Metabolomics Analysis of Liver from a Mouse Model of NPC1 Disease. J. Proteome Res. doi:10.1021/acs.jproteome.6b00238

Rustempasic, I., Can, M., (2013). Diagnosis of Parkinson's disease using Principal Component Analysis and Boosting Committee Machines. SOUTHEAST Eur. JOU RNAL O F SOFT Comput.

Salazar, D.A., Vélez, J.I., Salazar, J.C., (2012). Comparison between SVM and logistic regression: Which one is better to discriminate? Rev. Colomb. Estad. 35, 223–237.

Sandelowski, M., (2000). Focus on research methods combining qualitative and quantitative sampling, data collection, and analysis techniques. Res. Nurs. Health 23, 246–255.

Sanderson, K., (2015). Ebola investment boosts neglected-disease research. Nature. doi:10.1038/nature.(2015).18975

Sayre, N.L., Rimkunas, V.M., Graham, M.J., Crooke, R.M., Liscum, L., (2010). Recovery from liver disease in a Niemann-Pick type C mouse model. J. Lipid Res. 51, 2372–2383. doi:10.1194/jlr.M007211

Scannell, J.W., Blanckley, A., Boldon, H., Warrington, B., (2012). Diagnosing the decline in pharmaceutical R&D efficiency. Nat. Rev. Drug Discov. 11, 191–200. doi:10.1038/nrd3681

Schneider, J., 1997. Cross Validation [WWW Document]. URL https://www.cs.cmu.edu/~schneide/tut5/node42.html (accessed 1.3.17).

Schoonjans, F., (2016). Logistic regression [WWW Document]. MedCalc. URL https://www.medcalc.org/manual/logistic_regression.php (accessed 7.21.16).

Schripsema, J., (2010). Application of NMR in plant metabolomics: techniques, problems and prospects. Phytochem. Anal. 21, 14–21. doi:10.1002/pca.1185

Scikit Learn, (2016). 3.1. Cross-validation: evaluating estimator performance — scikit-learn 0.18.1 documentation [WWW Document]. URL http://scikit-learn.org/stable/modules/cross_validation.html (accessed 1.3.17).

SensaPharm, (2010). Anti-Malaria Drug Testing: Tests, Reference Standards and Assays.

Sharma, N., Sharma, P., Irwin, D., Shenoy, P., (2011). Predicting solar generation from weather forecasts using machine learning, in: Smart Grid Communications (SmartGridComm), (2011) IEEE International Conference On. IEEE, pp. 528–533.

Shen, Y., Zhu, J., (2009). Power analysis of principal components regression in genetic association studies. J. Zhejiang Univ. Sci. B 10, 721–730. doi:10.1631/jzus.B0830866

Shiryaeva, L., (2006). Proteomics and metabolomics in biological and medical applications. Planta 231, 1237–1249.

Shulaev, V., (2006). Metabolomics technology and bioinformatics. Brief. Bioinform. 7, 128–139. doi:10.1093/bib/bbl012

Siddiqui, N., (2003). Multicomponent analysis of encapsulated marine oil supplements using high-resolution 1H and 13C NMR techniques. J. Lipid Res. 44, 2406–2427. doi:10.1194/jlr.D300017-JLR200

Silva, D.P., (2002). The chemical logic behind... Gluconeogenesis [WWW Document]. URL http://homepage.ufp.pt/pedros/bq/gng.htm (accessed 6.8.17).

Smith, B.T., (2015). Implementing Logistic Regression From Scratch – Part 2: Python Code. Bryan Travis Smith PhD.

SMPDB, 2015a. Valine, Leucine, and Isoleucine Degradation [WWW Document]. URL http://smpdb.ca/view/SMP00032 (accessed 6.9.17).

SMPDB, 2015b. The Methionine Metabolism [WWW Document]. URL http://smpdb.ca/view/SMP00033 (accessed 6.11.17).

SMPDB, 2015c. The Lysine Degradation [WWW Document]. URL http://smpdb.ca/view/SMP00037 (accessed 6.11.17).

SMPDB, 2015d. The Glutamate Metabolism [WWW Document]. URL http://smpdb.ca/view/SMP00072 (accessed 6.12.17).

SMPDB, 2015e. The Alanine Metabolism [WWW Document]. URL http://smpdb.ca/view/SMP00055 (accessed 6.12.17).

SMPDB, 2015f. The Aspartate Metabolism [WWW Document]. URL http://smpdb.ca/view/SMP00067 (accessed 6.12.17).

SMPDB, 2015g. The Phenylalanine and Tyrosine Metabolism [WWW Document]. URL http://smpdb.ca/view/SMP00008 (accessed 6.12.17).

SMPDB, 2015h. The Taurine/Hypotaurine Metabolism [WWW Document]. URL http://smpdb.ca/view/SMP00021 (accessed 6.12.17).

SMPDB, (2010). The Cysteine Metabolism [WWW Document]. URL http://smpdb.ca/view/SMP00021 (accessed 5.27.17).

Softpedia, (2014). ACD/NMR Processor Academic Edition Download [WWW Document]. Softpedia. URL http://www.softpedia.com/get/Science-CAD/ACD-NMR-Processor-Academic-Edition.shtml (accessed 10.18.16).

Somorjai, R.L., Alexander, M.E., Baumgartner, R., Booth, S., Bowman, C., Demko, A., Dolenko, B., Mandelzweig, M., Nikulin, A.E., Pizzi, N.J., Pranckeviciene, E., Summers, A.R., Zhilkin, P., (2004). A Data-Driven, Flexible Machine Learning Strategy for the Classification of Biomedical Data, in: Artificial Intelligence Methods and Tools for System Biology. Springer, Netherland, pp. 1–231.

Sun, H., Zhang, A., Yan, G., Piao, C., Li, W., Sun, C., Wu, X., Li, X., Chen, Y., Wang, X., (2013). Metabolomic Analysis of Key Regulatory Metabolites in Hepatitis C Virus-infected Tree Shrews. Mol. Cell. Proteomics 12, 710–719. doi:10.1074/mcp.M112.019141

Sun, H., Zhang, A., Zou, D., Sun, W., Wu, X., Wang, X., (2014). Metabolomics Coupled with Pattern Recognition and Pathway Analysis on Potential Biomarkers in Liver Injury and Hepatoprotective Effects of Yinchenhao. Appl. Biochem. Biotechnol. 173, 857–869. doi:10.1007/s12010-014-0903-5

Tang, Y., Zhang, Y.-Q., Chawla, N.V., Krasser, S., (2009). SVMs modelling for highly imbalanced classification. IEEE Trans. Syst. Man Cybern. Part B Cybern. 39, 281–288.

Taooka, Y., Takezawa, G., Ohe, M., Sutani, A., Isobe, T., (2014). Multiple logistic regression analysis of risk factors in elderly pneumonia patients: QTc interval prolongation as a prognostic factor. Multidiscip. Respir. Med. 9, 1.

Tomita, M., Nishioka, T., (2006). Metabolomics: The Frontier of Systems Biology - M. Tomita, T. Nishioka - Google Books. Springer Science & Business Media.

Turkoglu, O., Zeb, A., Graham, S., Szyperski, T., Szender, J.B., Odunsi, K., Bahado-Singh, R., (2016). Metabolomics of biomarker discovery in ovarian cancer: a systematic review of the current literature. Metabolomics 12. doi:10.1007/s11306-016-0990-0

Tzoulaki, I., Ebbels, T.M.D., Valdes, A., Elliott, P., Ioannidis, J.P.A., (2014). Design and Analysis of Metabolomics Studies in Epidemiologic Research: A Primer on -Omic Technologies. Am. J. Epidemiol. 180, 129–139. doi:10.1093/aje/kwu143

van Gulik, W.M., (2010). Fast sampling for quantitative microbial metabolomics. Curr. Opin. Biotechnol. 21, 27–34. doi:10.1016/j.copbio.(2010).01.008

van Wyk, B., (2012). Research design and methods Part I.

Vanier, M.T., (2010). Orphanet Journal of Rare Diseases. Orphanet J. Rare Dis. 5, 16.

Vaus, D.A.D., Vaus, P.D. de, (2001). Research Design in Social Research. SAGE.

Venkatraman, N., (1988). The Concept of fit in Strategy Research: Towards Verbal and Statistical Correspondence.

Verplancke, T., Van Looy, S., Benoit, D., Vansteelandt, S., Depuydt, P., De Turck, F., Decruyenaere, J., (2008). Support vector machine versus logistic regression modelling for

prediction of hospital mortality in critically ill patients with haematological malignancies. BMC Med. Inform. Decis. Mak. 8. doi:10.1186/1472-6947-8-56

Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., 1999. Self-organizing map in Matlab: the SOM Toolbox, in: Matlab DSP. Presented at the 1999 Espoo, pp. 1607–3540.

Vishwanathan, S.V.M., Murty, M.N., (2002). SSVM: a simple SVM algorithm, in: Neural Networks, (2002). IJCNN'02. Proceedings of the (2002) International Joint Conference On. IEEE, pp. 2393–2398.

Vuckovic, D., (2012). Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry. Anal. Bioanal. Chem. 403, 1523–1548.

Wahyuni, D., (2012). The research design maze: Understanding paradigms, cases, methods and methodologies. J. Appl. Manag. Account. Res. 10, 69–80.

Wang, F., (2008). Biomarker Methods in Drug Discovery and Development. Springer Science & Business Media.

Wang, L., Jiang, M., Liao, S., Lu, Y., (2006). A feature selection method based on concept extraction and SOM text clustering analysis. Int. J. Comput. Sci. Netw. Secur. 6, 20–28.

Wang, X., Zhang, A., Han, Y., Wang, P., Sun, H., Song, G., Dong, T., Yuan, Y., Yuan, X., Zhang, M., Xie, N., Zhang, H., Dong, H., Dong, W., (2012). Urine Metabolomics Analysis for Biomarker Discovery and Detection of Jaundice Syndrome in Patients With Liver Disease. Mol. Cell. Proteomics 11, 370–380. doi:10.1074/mcp.M111.016006

Wang, X., Zhang, A., Sun, H., (2013). Power of metabolomics in diagnosis and biomarker discovery of hepatocellular carcinoma. Hepatology 57, 2072–2077. doi:10.1002/hep.26130

Wang, X.-Z., Dong, L.-C., Yan, J.-H., (2012). Maximum Ambiguity-Based Sample Selection in Fuzzy Decision Tree Induction. IEEE Trans. Knowl. Data Eng. 24, 1491–1505. doi:10.1109/TKDE.(2011).67

Wang, Y., Lawler, D., Larson, B., Ramadan, Z., Kochhar, S., Holmes, E., Nicholson, J.K., (2007). Metabonomic Investigations of Aging and Caloric Restriction in a Life-Long Dog Study. J. Proteome Res. 6, 1846–1854. doi:10.1021/pr060685n

Wang Y, Tetko IV, Hall MA, Frank E, Facius A, Mayer KFX, Mewes HW. Gene selection from microarray data for cancer classification—a machine learning approach. Comput Biol Chem. (2005);29(1):37–46. doi: 10.1016/j.compbiolchem.(2004).11.001

Wellman-Labadie, O., Zhou, Y., (2010). The US Orphan Drug Act: Rare disease research stimulator or commercial opportunity? Health Policy 95, 216–228. doi:10.1016/j.healthpol.(2009).12.001

Wijburg, F.A., Sedel, F., Pineda, M., Hendriksz, C.J., Fahey, M., Walterfang, M., Patterson, M.C., Iraith, J.E., Kolb, S.A., (2012). Development of a Suspicion Index to aid diagnosis of Niemann-Pick disease type C. Neurology 78, 1560–1567. doi:10.1212/WNL.0b013e3182563b82

Wizemann, T., Robison, S., Giffin, R., (2009). Drugs Development for Rare and Neglected Diseases and Individualised Therapies. The National Academies Press,

Wongravee, K., Lloyd, G.R., Silwood, C.J., Grootveld, M., Brereton, R.G., (2010). Supervised Self Organizing Maps for Classification and Determination of Potentially Discriminatory Variables: Illustrated by Application to Nuclear Magnetic Resonance Metabolomic Profiling. Anal. Chem. 82, 628–638. doi:10.1021/ac9020566

Woodford, C., Phillips, C., (2011). Numerical Methods with Worked Examples: Matlab Edition. Springer Science & Business Media.

World Health Organisation, (2014). World Malaria Report (2014). Geneva, Switzerland.

World Wide Antimalarial Resistance Network, (2012). Statistical Analysis Plan DHA-PQP Dose Impact Study Group.

Worley, B., Powers, R., (2015). Generalized adaptive intelligent binning of multiway data. Chemom. Intell. Lab. Syst. 146, 42–46. doi:10.1016/j.chemolab.(2015).05.005

Xi, Y., Rocke, D.M., (2008). Baseline Correction for NMR Spectroscopic Metabolomics Data Analysis. BMC Bioinformatics 9, 324. doi:10.1186/1471-2105-9-324

Xia, J., Broadhurst, D.I., Wilson, M., Wishart, D.S., (2013). Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics 9, 280–299. doi:10.1007/s11306-012-0482-9

XLSTAT, (2010). What is Principal Component Analysis (PCA)?

Xu, S., Zhou, S., Xia, D., Xia, J., Chen, G., Duan, S., Luo, J., (2010). Defects of synaptic vesicle turnover at excitatory and inhibitory synapses in Niemann–Pick C1-deficient neurons. Neuroscience 167, 608–620. doi:10.1016/j.neuroscience.(2010).02.033

Yang, C.-C., (2005). Six novel NPC1 mutations in Chinese patients with Niemann-Pick disease type C. J. Neurol. Neurosurg. Psychiatry 76, 592–595. doi:10.1136/jnnp.(2004).046045

Yin, H., (2002). Data visualisation and manifold mapping using the ViSOM. Neural Netw. 15, 1005–1016.

Yun, Y.-H., Deng, B.-C., Cao, D.-S., Wang, W.-T., Liang, Y.-Z., (2016). Variable importance analysis based on rank aggregation with applications in metabolomics for biomarker discovery. Anal. Chim. Acta. doi:10.1016/j.aca.(2015).12.043

Zhang, A., Sun, H., Wang, X., (2014). Urinary metabolic profiling of rat models revealed protective function of scoparone against alcohol induced hepatotoxicity. Sci. Rep. 4, 6768. doi:10.1038/srep06768

Zhang, X., Lu, X., Shi, Q., Xu, X., Hon-chiu, E.L., Harris, L.N., Iglehart, J.D., Miron, A., Liu, J.S., Wong, W.H., (2006). Recursive SVM feature selection and sample classification for mass-spectroscopy and microarray data. BMC Bioinformatics 7, 1.

ADDIN ZOTERO_BIBL {"custom":[]} CSL_BIBLIOGRAPHY

APPENDICES

1. DATA ANALYSIS

1.1. Factor Loadings

Chemical Shifts	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	12	F13
[0.68 .. 0.71]	-0.105	**0.805**	0.272	**-0.343**	0.174	0.105	-0.081	0.051	-0.073	0.007	0.022	-0.103	-0.086
[0.71 .. 0.73]	0.052	**0.909**	0.163	-0.244	0.147	0.064	-0.063	0.008	-0.078	0.034	0.061	-0.057	-0.013
[0.73 .. 0.75]	0.141	**0.925**	0.259	-0.152	0.059	0.027	-0.043	0.025	-0.043	0.017	0.025	-0.05	0.004
[0.75 .. 0.77]	0.159	**0.922**	0.277	-0.138	0.036	0.007	-0.025	0.016	-0.05	0.021	-0.003	-0.055	0
[0.77 .. 0.79]	0.114	**0.941**	0.154	-0.169	0.111	-0.026	-0.053	0.002	-0.044	0.023	-0.01	-0.046	-0.008
[0.81-0.89]	0.037	0.161	**0.581**	**0.601**	-0.172	-0.257	0	0.114	-0.124	-0.042	**-0.195**	-0.042	0.057
[0.89 .. 0.95]	-0.087	-0.056	**-0.61**	0.078	**0.629**	-0.174	**0**	-0.074	0.046	0.16	0.148	-0.006	-0.013
[0.95-1.06]	**0.555**	0.243	**-0.312**	**0.26**	0.03	0.085	-0.047	**-0.375**	0.048	0.306	0.15	-0.104	-0.089
[1.13 .. 1.15]	0.138	**0.563**	**0.3**	**-0.302**	**0.334**	0.159	**0.162**	0.086	-0.104	0.022	**0.272**	-0.137	-0.021
[1.15 .. 1.17]	0.137	**0.865**	0.186	-0.162	0.246	0.061	0.016	-0.02	0.092	0.024	0.07	0.005	-0.006
[1.17 .. 1.19]	0.075	0.555	0.074	0.052	0.113	0.136	-0.082	-0.001	-0.013	-0.327	0.131	0.451	-0.126
[1.19 .. 1.21]	0.3	0.59	0.261	0.237	0.006	-0.012	-0.039	0.011	0.157	-0.251	**0.269**	0.324	-0.032
[1.21 - 1.31]	**-0.76**	**-0.39**	**0.309**	0.126	**0.279**	-0.007	-0.018	-0.02	0.092	-0.031	0.128	-0.076	-0.062
[1.31 .. 1.37]	0.142	0.167	**-0.548**	**-0.657**	**-0.178**	0.016	**0.143**	**-0.151**	-0.118	0.015	-0.13	0.098	0.106
[1.45 .. 1.50]	**0.692**	-0.152	-0.15	0.055	-0.198	0.142	**0.26**	-0.093	0.012	0.334	0.083	0.08	0.168
[1.53 .. 1.55]	-0.207	**-0.741**	0.041	0.004	**0.377**	0.175	0.031	-0.031	0.107	0.047	0.221	0.084	-0.086
[1.55 .. 1.61]	-0.575	**-0.604**	-0.088	-0.095	**0.363**	0.132	-0.102	-0.032	0.133	-0.038	0.163	-0.065	-0.163
[1.69 .. 1.75]	**0.901**	0.035	-0.239	0.087	0.013	-0.044	-0.044	-0.063	0.124	0.169	-0.073	-0.039	-0.028

[1.80 .. 1.86]	0.335	-0.101	-0.199	0.186	**-0.588**	**0.291**	**-0.339**	0.058	0.03	-0.018	0.11	-0.161	-0.212
[1.90 .. 1.92]	0.411	**0.427**	**-0.49**	0.028	0.291	-0.107	-0.189	-0.009	-0.203	0.066	-0.124	-0.054	-0.04
[1.92 .. 1.94]	**0.673**	-0.092	0.269	-0.337	0.108	**0.312**	0	-0.232	0.187	0.135	-0.075	-0.005	-0.076
[1.94 .. 1.96]	0.595	0.069	**0.504**	0.214	0.232	0.066	-0.007	-0.183	0.243	0.043	-0.086	0.037	**0.247**
[1.96 .. 1.98]	0.365	0.062	**0.688**	**0.364**	0.071	-0.04	**0.239**	0.024	0.138	0.093	-0.052	0.025	**0.215**
[1.98 .. 2.03]	**-0.412**	**-0.345**	**0.641**	0.098	0.103	-0.09	0.03	**0.16**	0.166	0.091	**-0.232**	**-0.123**	-0.112
[2.03 .. 2.09]	0.504	0.162	-0.242	**-0.24**	**0.159**	-0.321	**-0.282**	**0.276**	**0.412**	0.076	**-0.147**	0	**-0.195**
[2.09 .. 2.12]	0.574	0.161	**-0.442**	0.095	**0.381**	-0.039	**-0.232**	0.221	0.22	-0.029	-0.089	-0.005	0.142
[2.12 .. 2.17]	**0.744**	0.12	0.124	0.084	**-0.275**	0.332	**-0.207**	-0.068	**0.289**	0.011	-0.002	-0.099	-0.12
[2.34 .. 2.39]	0.386	**-0.3**	-0.07	**-0.506**	**-0.411**	0.238	0.16	0.121	0.139	0.063	0.088	0.08	-0.016
[2.43 .. 2.48]	**0.791**	-0.055	-0.057	0.232	-0.142	**0.247**	-0.109	0.055	0.124	-0.104	-0.024	-0.019	**0.206**
[2.48 .. 2.50]	**0.618**	**-0.345**	-0.239	0.112	**0.304**	**0.355**	-0.011	**0.21**	0.019	-0.149	-0.173	-0.061	**0.21**
[2.50 .. 2.52]	**0.645**	**-0.33**	-0.088	0.036	0.243	**0.457**	0.108	**0.228**	-0.054	-0.14	-0.165	-0.068	-0.055
[2.52 .. 2.58]	**0.73**	**0.497**	**-0.239**	0.125	-0.153	-0.024	-0.006	**0.158**	-0.094	-0.081	0.037	0.041	0.027
[2.68 .. 2.74]	**0.737**	**0.313**	-0.113	0.066	**-0.295**	-0.034	-0.012	**0.321**	-0.066	-0.05	0.001	0.132	-0.085
[2.85 .. 2.88]	0.26	0.108	**0.254**	**0.601**	**0.217**	0.13	-0.079	**-0.194**	**-0.265**	0.18	**-0.255**	0.124	-0.24
[2.90 .. 2.94]	**0.706**	0.174	0.069	**0.423**	0.135	**0.271**	-0.059	-0.113	-0.228	0.05	-0.157	0.075	-0.183
[5.22 .. 5.24]	0.098	0.133	0.039	**0.786**	0.014	-0.013	-0.201	0.115	0.066	0.108	**0.336**	-0.01	0.113
[5.26 .. 5.37]	**-0.638**	**-0.375**	0.196	0.372	-0.177	0.062	**-0.153**	0.175	-0.005	0.099	0.124	0.045	-0.085
[6.87 .. 6.89]	0.429	**-0.654**	-0.244	0.159	0.123	0.008	0.013	0.052	**-0.258**	-0.082	0.097	0.15	-0.113
[6.89 .. 6.92]	**0.747**	-0.017	-0.195	0.231	-0.113	-0.212	0.158	0.041	-0.156	0.17	**0.296**	-0.096	0.102
[7.02 .. 7.08]	**0.852**	-0.115	-0.081	0.123	0.019	-0.013	**0.282**	-0.046	-0.05	-0.103	0.078	0.043	-0.003
[7.08 .. 7.14]	**0.791**	**-0.358**	0.22	-0.192	-0.118	-0.187	-0.11	-0.106	0.037	-0.067	-0.047	0.041	-0.088
[7.14 .. 7.16]	**0.696**	**-0.451**	0.094	-0.127	0.127	-0.168	**-0.239**	-0.136	-0.055	-0.122	0.017	0.098	0.143
[7.16 .. 7.18]	**0.738**	**-0.44**	-0.05	0.06	**0.28**	-0.069	-0.1	0.055	-0.083	-0.116	-0.038	-0.046	0.094
[7.18 .. 7.24]	**0.875**	0.053	0.021	0.063	-0.123	**-0.296**	0.077	0.097	0.006	0.127	0.103	-0.112	-0.034

[7.29 .. 7.35]	**0.886**	0.04	0.048	-0.094	-0.083	**-0.337**	-0.001	0.057	0.124	-0.039	-0.006	-0.024	-0.142
[7.35 .. 7.40]	**0.869**	-0.108	0.114	-0.058	0.048	**-0.265**	0.068	-0.088	-0.014	-0.136	0.06	-0.056	-0.184
[7.40 .. 7.45]	**0.704**	-0.313	0.367	-0.096	-0.089	-0.032	0.133	**-0.233**	0.027	-0.049	0.083	-0.088	-0.215
[7.45 .. 7.51]	0.513	**-0.645**	0.257	-0.132	0.188	**-0.229**	0.102	-0.016	-0.057	-0.112	0.039	-0.017	-0.042
[7.51 .. 7.56]	0.479	**-0.629**	0.232	-0.219	0.143	-0.011	-0.081	-0.02	**-0.252**	-0.008	0.026	0.026	0.001
[7.74 .. 7.79]	0.579	0.156	-0.144	0.108	**0.318**	0.031	**0.576**	0.141	-0.02	-0.099	-0.016	-0.168	-0.15
[7.79 .. 7.85]	0.559	-0.205	0.311	**-0.375**	0.032	-0.074	**-0.22**	**-0.322**	0.053	-0.043	0.009	0.184	0.181
[7.85 .. 7.91]	0.467	**-0.416**	0.391	**-0.291**	0.064	0.097	0.072	**0.257**	0.039	-0.041	0.205	-0.128	0.018
[7.91 .. 7.97]	0.323	**-0.475**	**0.406**	**-0.357**	-0.133	-0.039	**-0.323**	0.094	-0.195	0.068	-0.043	-0.026	-0.063
[8.17 .. 8.23]	0.08	-0.204	0.188	**-0.273**	0.148	0.012	0.131	**0.375**	0.008	0.545	-0.05	0.461	-0.084
[8.28 .. 8.31]	0.304	-0.167	**0.395**	-0.332	0.049	0.123	**-0.411**	0.18	**-0.359**	0.143	0.144	-0.134	0.182

Table 47. Full visualisation of contributions of the NPC1 disease features to each one of the factors where positive values are showing important contribution to the corresponding factor.

1.2. Factor Scores

Observation	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
Obs1	-0.566	-2.816	-0.299	-2.194	1.548	1.377	0.373	0.482	0.969	3.066	0.130	2.107	-0.179
Obs2	1.792	-1.611	-4.272	-1.779	1.358	-1.701	0.793	-0.774	-1.115	0.139	-0.635	0.064	0.300
Obs3	0.066	-2.933	0.119	-1.895	-0.294	0.511	0.338	-1.603	-0.846	-0.240	0.061	0.274	0.433
Obs4	1.057	-2.154	0.179	0.445	0.672	-0.716	0.410	-0.769	-0.021	0.429	0.247	-0.703	-0.343
Obs5	-1.680	-2.818	1.183	-1.070	-0.157	1.470	1.023	-0.868	-0.127	-0.245	-0.102	-0.072	0.813
Obs6	-4.669	-0.289	0.149	-1.912	-0.869	-0.352	0.539	-2.523	-0.295	0.011	0.112	-0.032	0.819

Obs7	-3.849	2.158	-1.911	-1.626	-0.568	1.353	0.572	-0.095	0.330	-0.868	-0.539	-0.412	0.799
Obs8	-0.968	-0.608	1.132	-0.445	0.566	-0.966	-0.216	-0.077	0.457	-0.596	-0.870	-0.598	-0.001
Obs9	-0.051	-3.845	-2.928	**2.824**	**2.408**	1.418	-0.980	0.624	1.849	1.208	-0.274	-0.010	1.374
Obs10	**3.259**	-0.402	0.441	-2.988	-1.142	0.190	-1.006	-2.655	1.320	0.317	0.306	0.879	1.473
Obs11	**3.219**	-2.337	-0.646	-0.499	0.566	0.520	1.008	0.632	1.080	-0.727	0.854	-0.579	-0.084
Obs12	0.672	2.872	-0.215	-0.977	-0.114	-0.083	1.000	0.384	1.333	-1.127	-0.306	0.123	1.435
Obs13	**3.608**	-0.014	0.425	-1.608	-1.329	2.132	0.432	-1.138	0.893	-0.517	-2.035	0.709	1.766
Obs14	-11.125	-1.433	1.803	0.457	1.213	-1.408	0.264	0.683	-0.291	-0.501	0.713	-0.502	0.228
Obs15	-12.453	-4.388	-0.176	-0.416	0.680	-1.281	0.172	0.036	-0.928	-0.792	0.064	-1.616	-0.463
Obs16	-3.794	-3.083	0.957	-0.562	-0.883	0.231	1.478	0.532	-0.588	0.059	-0.986	-0.737	0.373
Obs17	-3.565	-0.499	-0.522	0.411	**2.074**	2.355	0.017	-0.354	1.098	0.115	1.071	-0.188	0.569
Obs18	-3.927	-2.328	0.587	-1.094	0.335	0.393	-1.224	-2.325	0.579	0.307	0.511	0.613	1.715
Obs19	-3.446	-0.922	1.579	-0.749	**2.146**	1.834	0.002	-0.791	1.331	-0.188	1.719	0.761	-0.164
Obs20	**7.482**	-2.717	1.297	-3.045	0.713	-2.429	-1.659	-1.023	0.701	-1.117	0.200	1.124	-0.799
Obs21	**6.387**	-2.061	1.874	-3.236	-0.649	3.710	-4.748	-0.409	-0.980	1.168	0.185	-0.655	1.437
Obs22	-2.624	-2.858	-1.138	0.293	0.800	-0.290	-0.388	-0.660	0.580	-0.671	0.205	-0.494	-0.521
Obs23	-6.253	-2.913	-2.265	-1.005	-1.662	1.662	-0.610	-0.231	-0.785	-0.765	-0.174	-1.100	-1.433
Obs24	-5.780	-3.902	-1.921	-0.922	1.128	0.811	-0.198	-0.711	-0.620	-0.640	0.501	-0.782	-0.392
Obs25	-2.798	1.793	-2.981	-0.970	**2.850**	1.325	-0.883	-0.959	0.361	0.398	0.808	-0.506	0.010
Obs26	-5.864	-2.661	-0.206	-2.156	1.177	1.182	-1.393	-1.644	0.144	-0.333	1.036	0.368	-0.148
Obs27	1.220	0.297	-3.029	-1.198	-2.199	1.848	-1.885	0.299	2.121	-0.898	-0.213	-1.271	-3.064
Obs28	-2.918	-0.826	-0.908	-1.817	1.063	2.158	0.265	-0.127	0.384	-0.302	0.276	-0.252	0.113
Obs29	-4.674	**2.509**	1.461	1.291	-1.480	-0.630	0.351	0.893	1.142	-0.216	-0.504	-1.112	0.413
Obs30	-0.454	-5.158	**3.103**	-1.853	-2.331	0.063	-3.943	0.923	-0.706	0.221	-0.246	-1.211	0.163
Obs31	-3.829	**6.068**	**3.455**	-1.960	-2.544	-1.254	-1.120	-1.446	-0.014	-1.455	0.133	1.776	0.165
Obs32	-1.291	-2.322	-0.855	-0.136	-0.900	-1.299	0.172	-0.346	-0.057	-0.164	-0.778	-0.678	-0.799
Obs33	-5.996	**4.405**	1.550	-2.557	-1.203	-0.400	1.472	2.802	-0.707	0.902	-0.950	1.015	-0.839
Obs34	-5.038	-5.770	0.250	-1.147	-0.320	-2.449	-0.107	1.344	-1.214	0.202	-0.845	-1.257	-0.917
Obs35	-3.538	-2.548	1.076	0.512	1.017	-0.384	-0.845	0.929	0.505	-0.849	0.107	-0.772	-0.759

Obs36	-2.500	-2.587	-0.289	-2.779	-0.332	0.010	-0.164	2.014	-0.117	-0.938	-0.787	0.418	-1.231
Obs37	**3.295**	-3.092	0.571	-0.163	-0.840	-1.170	0.599	-0.260	-1.026	-0.771	-0.239	-0.293	-1.060
Obs38	-1.568	-2.820	1.719	-0.301	0.688	0.877	-0.049	1.184	0.908	-1.473	-1.277	-0.676	0.512
Obs39	-1.503	1.212	-5.691	-0.845	**4.120**	1.138	-1.845	2.744	-0.218	-2.241	-2.320	0.252	0.986
Obs40	-7.016	**4.413**	**2.261**	-2.468	-1.045	0.153	-0.088	-0.303	-0.280	-0.753	-1.037	-0.255	-0.001
Obs41	**3.452**	-0.754	-2.237	0.292	**2.527**	-1.104	-0.159	0.317	-0.565	-1.185	0.306	1.315	1.455
Obs42	**2.710**	-2.338	-0.987	-1.582	-1.152	0.146	1.624	-0.552	0.673	-0.658	-1.059	0.207	0.709
Obs43	-4.235	0.644	-3.092	0.402	1.940	0.939	-0.197	0.269	0.074	-1.598	2.186	1.520	-0.188
Obs44	-4.853	-0.112	0.411	1.439	0.361	0.293	-0.058	0.427	0.533	-0.724	0.409	-0.963	0.765
Obs45	1.580	-2.786	0.188	-3.098	1.534	-0.130	2.076	0.844	-0.796	1.010	1.230	-2.046	-2.162
Obs46	**3.523**	**6.420**	-3.585	-0.158	-1.403	-2.785	0.179	-0.658	0.091	0.800	1.166	-0.832	1.113
Obs47	**2.320**	-2.766	**2.549**	-1.914	0.406	-2.344	-2.402	1.217	1.295	0.043	-0.270	2.792	-0.313
Obs48	-0.506	**4.837**	-3.749	-0.347	-0.421	-0.712	-1.047	0.600	-0.445	0.241	-1.330	-1.015	0.137
Obs49	-1.552	-2.354	3.774	-0.757	-0.826	-0.336	-0.942	-0.739	1.779	-0.356	-1.141	0.941	1.900
Obs50	-2.111	-1.789	-1.546	-1.397	1.027	1.138	1.783	-1.203	-0.841	0.079	0.675	-0.594	0.083
Obs51	**7.090**	**5.593**	**6.890**	-5.407	**2.581**	2.031	1.378	2.128	-3.908	-0.647	4.000	-1.422	1.062
Obs52	-2.843	**7.181**	0.922	-1.600	-2.393	-0.578	0.807	-0.629	-0.009	-0.483	-0.505	-0.064	0.047
Obs53	1.387	-0.993	-1.950	1.498	1.199	-2.000	1.198	-1.372	1.322	1.390	0.664	-1.253	0.355
Obs54	-8.427	**2.780**	1.239	-1.125	0.920	0.527	0.033	-1.320	0.975	-0.374	2.028	-0.081	-0.641
Obs55	-0.123	**3.855**	-0.163	-0.818	-1.993	1.995	1.049	-0.261	0.348	-0.595	0.436	-1.272	0.170
Obs56	-1.797	-2.181	-1.705	0.433	-0.359	1.520	-0.263	0.139	0.840	-0.219	-0.853	-0.510	-0.420
Obs57	-4.273	**3.501**	1.407	-1.563	-2.204	1.225	1.738	-0.460	-0.824	-1.215	-0.191	-0.707	2.083
Obs58	0.471	-0.928	-0.463	-0.198	-1.323	2.194	1.037	0.750	0.521	0.923	-2.174	-0.267	1.016
Obs59	**5.250**	-6.314	-0.981	-0.321	-1.847	-0.106	1.888	1.135	0.034	-0.829	0.187	-0.121	-0.250
Obs60	**2.848**	-2.122	-0.330	-2.794	1.158	0.587	1.277	1.411	-1.584	0.892	-0.019	1.167	0.696
Obs61	-3.222	-0.195	0.434	-0.323	0.901	-1.062	0.576	0.400	0.752	-1.114	0.618	-0.500	0.037
Obs62	**4.989**	**5.829**	**2.300**	1.801	-0.399	-0.386	0.854	1.381	1.849	-1.643	2.560	-0.382	-0.686
Obs63	**2.120**	-2.416	-2.135	-2.463	-0.389	-0.306	2.064	-0.117	-0.217	-0.684	0.393	-0.448	-0.043
Obs64	1.469	-2.599	-0.747	-1.337	1.937	-0.107	0.905	1.309	2.064	0.332	0.590	0.226	-0.476

Obs65	**6.535**	-1.822	-1.167	-2.522	-1.894	0.088	2.947	2.107	1.589	1.443	-0.460	1.979	-0.062
Obs66	-2.482	**2.063**	-0.245	-2.645	0.268	0.099	2.556	0.890	0.680	3.003	0.854	2.313	-0.161
Obs67	**7.898**	0.269	0.645	-2.576	1.364	-1.362	1.323	0.060	-0.401	-1.434	0.596	0.562	-0.053
Obs68	**6.920**	-5.245	-4.012	3.993	-0.943	1.058	0.313	0.937	1.530	-0.992	0.815	-0.990	1.137
Obs69	-3.540	-6.066	**5.364**	-1.762	-1.017	-1.172	-4.575	1.929	-1.773	1.094	-0.599	-0.946	1.000
Obs70	-1.691	-0.166	1.727	-1.733	-3.043	-2.147	-2.054	-0.398	0.851	0.566	0.808	-0.557	0.193
Obs71	-6.633	-1.525	1.476	-2.269	0.200	-0.014	2.413	1.637	0.479	3.647	0.360	1.558	-0.940
Obs72	-2.945	1.102	-1.970	0.862	**2.216**	-1.170	0.387	-1.858	-0.349	1.127	0.777	-0.500	0.605
Obs73	-3.837	1.798	-1.104	1.760	1.621	-1.386	-0.549	-0.885	0.413	0.112	0.912	0.801	1.290
Obs74	-3.026	-1.489	-6.520	0.092	**3.549**	-2.293	-1.497	-0.537	-2.029	1.444	0.830	0.396	-0.858
Obs75	0.798	-1.643	1.779	1.701	0.517	-2.515	-0.167	-1.397	-0.259	0.129	-0.383	0.503	-0.777
Obs76	0.686	**3.716**	-2.443	0.715	-0.458	-0.635	0.164	-0.683	-0.517	0.575	-1.388	-0.480	1.083
Obs77	1.637	-2.326	1.072	1.667	0.604	0.831	0.380	-1.637	-0.580	1.092	-0.413	-0.330	0.372
Obs78	0.385	0.049	1.544	**2.141**	-0.622	-1.706	1.244	0.529	0.452	-0.618	-0.512	-0.388	0.122
Obs79	-0.573	**4.622**	**2.259**	**2.979**	0.292	0.210	1.138	0.168	0.438	0.474	0.507	-0.280	1.282
Obs80	1.987	0.449	0.800	**2.590**	1.039	-0.290	0.358	0.787	0.624	-0.472	-0.447	0.350	1.613
Obs81	0.971	-2.024	1.928	**6.059**	1.740	-0.171	-0.638	0.913	0.949	0.042	1.619	-0.264	0.542
Obs82	-2.188	-1.838	1.927	**2.058**	0.617	-0.284	1.387	0.467	-0.283	-0.767	-0.487	-0.665	2.023
Obs83	1.206	-4.109	0.485	1.662	0.390	-1.290	0.955	1.382	0.693	-0.687	-0.545	-0.099	-0.224
Obs84	1.019	**3.197**	**2.539**	**6.107**	-1.843	1.241	-1.723	0.277	0.388	0.492	1.248	-0.228	-1.768
Obs85	**2.604**	**9.907**	-1.134	-0.683	**3.906**	-0.383	-1.927	-0.458	-1.440	0.668	-2.442	0.641	-1.548
Obs86	-2.852	**6.547**	1.545	0.808	0.411	1.055	-0.877	0.619	0.190	0.979	-0.414	0.598	-1.765
Obs87	-4.506	-0.481	**2.586**	1.344	1.364	-1.468	-0.257	0.278	0.263	0.524	0.086	0.308	-0.385
Obs88	0.833	-1.998	**3.665**	**2.022**	0.728	-0.863	0.536	1.290	-1.505	0.689	-0.703	1.329	0.586
Obs89	0.000	-1.714	1.724	**2.418**	-0.372	-1.126	1.281	-0.349	-0.599	-0.522	-1.538	0.202	0.223
Obs90	1.111	-0.887	0.852	**2.615**	0.874	-0.175	-0.026	-0.207	-0.481	1.334	0.026	-0.273	-0.815
Obs91	0.967	**3.228**	1.712	1.378	0.844	0.196	0.195	0.715	0.560	0.176	-1.238	-0.111	-0.208
Obs92	-1.245	1.799	**3.001**	1.620	-0.561	-0.438	0.011	-0.754	-0.407	-1.255	-1.580	0.999	-0.832
Obs93	1.461	-0.839	0.438	1.193	-0.626	-1.813	1.029	-2.569	-0.330	1.117	0.472	0.044	-0.693

Obs94	-4.881	**2.931**	0.939	**3.758**	0.084	2.619	-1.123	0.112	0.647	1.509	-0.560	0.282	0.352
Obs95	1.352	**2.297**	1.941	**3.253**	-0.611	0.461	1.063	-0.427	-0.930	0.192	-1.394	-0.031	0.187
Obs96	**2.218**	-0.332	0.201	1.289	1.728	1.383	-0.549	0.443	0.245	-0.252	-0.774	0.387	-0.452
Obs97	0.362	-1.267	1.787	**2.174**	-0.783	-0.131	1.519	-1.079	-1.455	-0.002	-0.449	-0.490	-0.345
Obs98	-0.369	-2.617	-0.708	1.578	-0.892	1.360	0.868	-1.927	-0.354	0.697	-0.607	-0.511	-0.366
Obs99	-2.501	-1.226	-1.239	0.912	1.053	0.988	-0.112	-1.099	-0.246	0.284	0.416	0.003	-0.603
Obs100	1.326	-2.880	0.324	1.256	-0.315	0.689	0.348	-0.976	-2.661	-0.129	-0.567	0.095	-1.077
Obs101	-5.267	-1.428	0.640	**2.061**	1.291	1.437	-0.435	-0.096	-0.903	-4.150	0.668	4.651	-1.345
Obs102	**2.186**	-4.083	1.695	**3.307**	0.842	-1.233	1.050	0.528	-0.109	0.004	-0.505	0.683	0.331
Obs103	**2.929**	**4.662**	**4.706**	**5.893**	1.701	-0.271	-0.709	1.006	0.773	0.348	1.182	-0.702	-0.668
Obs104	-2.179	**6.332**	0.368	-1.409	-2.194	0.525	0.820	-0.096	0.032	-0.280	0.146	-0.247	-0.403
Obs105	**7.353**	-0.934	-0.333	0.655	-0.684	0.883	1.497	0.789	-0.129	-1.073	0.659	0.221	-0.156
Obs106	**5.818**	-1.027	-0.331	-0.961	-2.532	-0.004	1.179	-0.284	-1.120	-0.754	0.399	-1.044	-0.458
Obs107	**6.577**	1.983	-4.193	0.172	0.702	-0.112	-1.102	0.118	-1.158	-1.751	0.087	1.251	-0.179
Obs108	**6.110**	**4.344**	-1.406	-1.326	**2.646**	-0.834	-0.916	1.082	-1.647	-0.047	-0.459	-0.906	0.703
Obs109	-4.175	**7.650**	-4.149	1.062	-2.013	-0.786	-1.817	2.628	-1.197	1.888	0.299	0.071	1.929
Obs110	-3.976	-4.837	-6.059	**4.460**	-9.104	0.271	-0.984	2.157	-2.318	0.665	3.339	2.038	0.554
Obs111	**5.530**	-1.352	-1.793	0.838	-2.999	3.078	-0.517	-0.052	-0.386	0.366	0.092	0.261	-1.198
Obs112	**2.315**	**5.792**	-1.446	-0.555	0.515	0.462	-0.742	0.187	-0.883	-0.753	0.520	-0.559	-0.802
Obs113	-1.522	**4.792**	-4.636	0.514	-1.106	-3.058	-0.975	1.184	-0.638	-0.558	0.925	0.479	1.393
Obs114	**6.920**	**2.357**	-2.088	0.017	0.133	0.513	-0.241	-0.239	-0.886	-0.636	-0.194	-0.778	-0.924
Obs115	1.706	**3.282**	-0.041	-1.439	-0.629	-0.322	-0.070	0.173	1.597	-0.470	0.133	-0.250	-0.746
Obs116	**10.793**	-2.694	1.714	-1.902	-1.262	-1.285	-2.934	-1.827	-0.078	0.190	1.522	0.619	0.735
Obs117	1.171	**3.624**	-2.596	-0.696	0.855	-1.504	-1.261	-0.032	1.596	0.701	-0.100	-1.372	-0.496
Obs118	-2.695	**6.371**	1.121	-1.429	-2.528	-0.592	0.446	-1.474	-0.505	0.042	-0.363	1.011	-1.372
Obs119	**7.662**	0.358	**2.796**	0.733	0.810	0.208	0.075	0.208	-1.937	1.205	0.023	-0.521	0.150
Obs120	**4.117**	-1.302	1.446	3.311	1.003	2.718	-1.073	-0.163	-1.452	0.536	-1.154	0.497	-1.152
Obs121	**2.309**	-0.672	-0.446	1.040	-0.424	-1.504	1.061	-1.944	-1.107	-0.098	-1.191	0.195	-0.409
Obs122	**4.130**	-0.469	-0.330	2.685	0.825	-0.107	0.167	0.301	-1.124	0.294	-0.919	-0.145	0.682

Obs123	**3.473**	**3.743**	-0.425	1.951	-0.697	0.989	0.218	-0.670	-0.309	-0.543	-1.226	0.732	-0.753
Obs124	0.087	**3.681**	-2.042	1.482	-2.360	1.342	0.268	-1.290	-0.307	1.664	-0.659	0.375	-0.069
Obs125	**2.582**	-1.002	-1.201	0.855	1.370	0.536	0.611	-1.898	-0.013	1.918	0.244	0.294	-0.593
Obs126	-1.493	-3.012	0.354	1.617	-1.431	-1.146	-0.327	0.308	0.185	-0.112	0.024	-0.947	-0.645
Obs127	**5.368**	**2.035**	**2.407**	0.496	-0.041	-0.459	-0.679	-1.581	2.095	0.595	1.101	-0.399	-0.044
Obs128	1.579	0.963	0.781	-3.004	1.309	0.649	-0.579	2.070	1.524	1.905	-0.496	-0.631	-0.343
Obs129	**4.565**	-3.745	-0.306	-3.249	-1.640	-1.176	-1.346	0.305	1.483	0.685	-0.520	0.389	-0.767
Obs130	**3.435**	**2.104**	-1.284	-1.520	-1.831	-1.210	-0.356	0.875	3.417	-0.597	-0.428	-0.886	-0.783

Table 48. Full visualisation of the factor scores related to the scores of the contribution of each observation to the factors or principal components of the NPC1 disease features. The high positive values expressing important contribution to the corresponding factor. For instance, observation 10 is making a big contribution to the factor F1.

1.3. Chemical shifts/buckets of biomolecules related to NPC1 disease dataset

Features Number	Chemical Shift	Molecules Names
1	[0.68 .. 0.71]	L-Lactate
2	[0.71 .. 0.73]	Noisy – Noisy
3	[0.73 .. 0.75]	16b-Hydroxyestradiol
4	[0.75 .. 0.77]	Noisy – Noisy
5	[0.77 .. 0.79]	Masoprocol
6	[0.81-0.89]	Hexacosanoate
7	[0.89 .. 0.95]	L-Isoleucine

8	[0.95-1.06]	Leucine; Valine
9	[1.13 .. 1.15]	Propionylglycine
10	[1.15 .. 1.17]	Ethanol
11	[1.17 .. 1.19]	Isopropyl hexadecanoate
12	[1.19 .. 1.21]	3-Aminoisobutyrate
13	[1.21 - 1.31]	(R)-3-Hydroxybutyrate; L-Fucose;
14	[1.31 .. 1.37]	Lactate; 3-Hydroxyisovalerate
15	[1.45 .. 1.50]	Alanine
16	[1.53 .. 1.55]	***Adipate***
17	[1.55 .. 1.61]	***Adipate***
18	[1.69 .. 1.75]	Lysine
19	[1.80 .. 1.86]	Thymine; 2-Hydroxyglutarate
20	[1.90 .. 1.92]	Lysine
21	[1.92 .. 1.94]	Acetic acid
22	[1.94 .. 1.96]	Ornithine
23	[1.96 .. 1.98]	Methylglutaric acid
24	[1.98 .. 2.03]	2-Hydroxyglutarate
25	[2.03 .. 2.09]	N-Acetyl-4-O-acetylneuraminic acid
26	[2.09 .. 2.12]	Glutamine
27	[2.12 .. 2.17]	Methionine; Glutamine
28	[2.34 .. 2.39]	Pyruvate; D-Glutamate
29	[2.43 .. 2.48]	Glutamine
30	[2.48 .. 2.50]	L-Glutamine

31	[2.50 .. 2.52]	Methylamine
32	[2.52 .. 2.58]	Citrate
33	[2.68 .. 2.74]	Citrate
34	[2.85 .. 2.88]	Trimethylamine
35	[2.90 .. 2.94]	Dimethylglycine
36	[5.22 .. 5.24]	α-Glucose
37	[5.26 .. 5.37]	α-Glucose; Allantoate
38	[6.87 .. 6.89]	3-(3-Hydroxyphenyl)-3-hydroxypropanoate
39	[6.89 .. 6.92]	3-hydrophenyl
40	[7.02 .. 7.08]	Phenol
41	[7.08 .. 7.14]	L-Histidine
42	[7.14 .. 7.16]	Syringic acid
43	[7.16 .. 7.18]	Tyrosine
44	[7.18 .. 7.24]	Indoxyl sulfate
45	[7.29 .. 7.35]	Methyl phenylacetate
46	[7.35 .. 7.40]	Indoxyl sulfate; Phenylalanine
47	[7.40 .. 7.45]	Indoxyl sulfate
48	[7.45 .. 7.51]	Indoxyl sulfate; Benzoate
49	[7.51 .. 7.56]	***Hippurate***
50	[7.74 .. 7.79]	4-Hydroxybenzoate
51	[7.79 .. 7.85]	7-Methylxanthine; 4-Hydroxybenzoate
52	[7.85 .. 7.91]	1-Methylhistidine
53	[7.91 .. 7.97]	Quinolinate
54	[8.17 .. 8.23]	Trigonelline
55	[8.28 .. 8.31]	Hypoxanthine

Table 49. Complete chemical shifts/buckets related to the NPC1 disease dataset, used to identify the different molecules involved in the ranking established by the different techniques employed in this research.

1.4. XLSTAT Generated Full Statistics Generated for NPC1 Dataset

XLSTAT 2016.07.39157 - Principal Component Analysis (PCA) - Start time: 26/12/2016 at 20:38:01
Observations/variables table: Workbook = PLASMA.1.03.xlsm / Sheet = Sheet2 / Range = Sheet2!
PCA type: Pearson (n)
Type of biplot: Correlation biplot / Coefficient = Automatic
Run again:

Summary statistics:

Variable	Observations	with missing	thout missi	Minimum	Maximum	Mean	Std. deviation
[0.68 .. 0.71]	130	0	130	0.000	0.003	0.001	0.000
[0.71 .. 0.73]	130	0	130	0.000	0.002	0.001	0.000
[0.73 .. 0.75]	130	0	130	0.000	0.002	0.001	0.000
[0.75 .. 0.77]	130	0	130	0.000	0.003	0.001	0.000
[0.77 .. 0.79]	130	0	130	0.000	0.004	0.002	0.001
[0.81-0.89]	130	0	130	0.095	0.187	0.132	0.018
[0.89 .. 0.95]	130	0	130	0.016	0.051	0.028	0.005
[0.95-1.06]	130	0	130	0.019	0.038	0.028	0.004
[1.13 .. 1.15]	130	0	130	0.000	0.006	0.002	0.001
[1.15 .. 1.17]	130	0	130	0.001	0.005	0.003	0.001
[1.17 .. 1.19]	130	0	130	0.004	0.018	0.006	0.001
[1.19 .. 1.21]	130	0	130	0.007	0.024	0.014	0.002
[1.21 - 1.31]	130	0	130	0.158	0.391	0.241	0.042
[1.31 .. 1.37]	130	0	130	0.062	0.284	0.203	0.036
[1.45 .. 1.50]	130	0	130	0.009	0.022	0.016	0.002
[1.53 .. 1.55]	130	0	130	0.001	0.006	0.003	0.001
[1.55 .. 1.61]	130	0	130	0.005	0.030	0.014	0.005
[1.69 .. 1.75]	130	0	130	0.006	0.015	0.011	0.002
[1.80 .. 1.86]	130	0	130	0.003	0.034	0.011	0.005
[1.90 .. 1.92]	130	0	130	0.002	0.014	0.004	0.002
[1.92 .. 1.94]	130	0	130	0.001	0.005	0.003	0.001
[1.94 .. 1.96]	130	0	130	0.003	0.006	0.005	0.001
[1.96 .. 1.98]	130	0	130	0.006	0.012	0.009	0.001
[1.98 .. 2.03]	130	0	130	0.024	0.046	0.035	0.004
[2.03 .. 2.09]	130	0	130	0.034	0.057	0.044	0.005
[2.09 .. 2.12]	130	0	130	0.003	0.009	0.005	0.001
[2.12 .. 2.17]	130	0	130	0.006	0.020	0.013	0.002
[2.34 .. 2.39]	130	0	130	0.003	0.025	0.011	0.004
[2.43 .. 2.48]	130	0	130	0.003	0.015	0.008	0.002
[2.48 .. 2.50]	130	0	130	0.000	0.003	0.001	0.001
[2.50 .. 2.52]	130	0	130	0.000	0.002	0.001	0.000
[2.52 .. 2.58]	130	0	130	0.023	0.080	0.055	0.012
[2.68 .. 2.74]	130	0	130	0.014	0.044	0.031	0.006
[2.85 .. 2.88]	130	0	130	0.000	0.014	0.003	0.003
[2.90 .. 2.94]	130	0	130	0.002	0.011	0.006	0.002
[5.22 .. 5.24]	130	0	130	0.000	0.009	0.002	0.002
[5.26 .. 5.37]	130	0	130	0.019	0.058	0.031	0.006
[6.87 .. 6.89]	130	0	130	0.000	0.000	0.000	0.000
[6.89 .. 6.92]	130	0	130	0.000	0.002	0.001	0.000
[7.02 .. 7.08]	130	0	130	0.000	0.002	0.001	0.000
[7.08 .. 7.14]	130	0	130	0.000	0.002	0.001	0.000
[7.14 .. 7.16]	130	0	130	0.000	0.001	0.000	0.000
[7.16 .. 7.18]	130	0	130	0.000	0.000	0.000	0.000
[7.18 .. 7.24]	130	0	130	0.001	0.003	0.002	0.000
[7.29 .. 7.35]	130	0	130	0.001	0.003	0.002	0.000
[7.35 .. 7.40]	130	0	130	0.000	0.001	0.001	0.000
[7.40 .. 7.45]	130	0	130	0.000	0.001	0.001	0.000
[7.45 .. 7.51]	130	0	130	0.000	0.000	0.000	0.000
[7.51 .. 7.56]	130	0	130	0.000	0.000	0.000	0.000
[7.74 .. 7.79]	130	0	130	0.000	0.002	0.001	0.000
[7.79 .. 7.85]	130	0	130	0.000	0.002	0.000	0.000
[7.85 .. 7.91]	130	0	130	0.000	0.001	0.000	0.000
[7.91 .. 7.97]	130	0	130	0.000	0.001	0.000	0.000
[8.17 .. 8.23]	130	0	130	0.000	0.001	0.000	0.000
[8.28 .. 8.31]	130	0	130	0.000	0.000	0.000	0.000

Table 50. Statistics Related to the NPC LDD, with the values of the means and standard deviation of the features provided.

1.5. NPC1 Disease Plasma Dataset

	L-Lactic ac	16b-Hydro	Masoproc	Hexacosar	L-Isoleucir	Leucine	Propionylg	Ethanol	Isopropyl I	3-Aminois	(R)-3-Hydr	3-Hydroxy	Alanine
S1	0.00111	0.000664	0.00144	0.113	0.0322	0.0277	0.00204	0.00338	0.00609	0.0112	0.26	0.211	0.021
S2	0.000953	0.000634	0.00148	0.106	0.0352	0.0304	0.00179	0.00299	0.00621	0.00961	0.197	0.284	0.0174
S3	0.000934	0.000608	0.00128	0.119	0.0266	0.0262	0.00204	0.00321	0.00564	0.0113	0.246	0.239	0.0169
S4	0.00104	0.000704	0.00135	0.135	0.0285	0.0332	0.00194	0.00316	0.00549	0.0129	0.257	0.191	0.0155
S5	0.00117	0.0007	0.00146	0.132	0.0262	0.0246	0.00198	0.00322	0.00561	0.0114	0.276	0.202	0.0182
S6	0.00128	0.000835	0.00162	0.121	0.0257	0.0262	0.00233	0.00333	0.00558	0.0113	0.27	0.244	0.0149
S7	0.00145	0.00102	0.00189	0.109	0.0259	0.0257	0.00249	0.00386	0.00649	0.0131	0.248	0.261	0.0131
S8	0.00135	0.000886	0.00171	0.145	0.028	0.0262	0.00216	0.0035	0.00598	0.0128	0.258	0.193	0.0139
S9	0.000434	0.000283	0.000846	0.115	0.0342	0.033	0.00173	0.00279	0.00471	0.0113	0.287	0.159	0.0158
S10	0.00136	0.00099	0.00187	0.114	0.0264	0.0314	0.00233	0.00379	0.00663	0.0131	0.217	0.238	0.0201
S11	0.00084	0.0006	0.00126	0.116	0.0268	0.0307	0.0032	0.00346	0.00605	0.0141	0.232	0.21	0.0162
S12	0.00141	0.00118	0.00227	0.13	0.0248	0.0272	0.00281	0.00424	0.00721	0.0156	0.224	0.22	0.0171
S13	0.00113	0.00104	0.00193	0.132	0.0233	0.0275	0.00205	0.0033	0.00568	0.0134	0.196	0.242	0.0188
S14	0.00117	0.000682	0.00132	0.134	0.0282	0.019	0.00208	0.00367	0.00558	0.0109	0.376	0.154	0.009
S15	0.00104	0.000289	0.000769	0.123	0.0285	0.0187	0.0015	0.00245	0.00393	0.00807	0.391	0.171	0.00915
S16	0.0011	0.000612	0.00132	0.145	0.0246	0.0236	0.00173	0.00275	0.00462	0.00974	0.278	0.192	0.0162
S17	0.00125	0.000785	0.00167	0.112	0.0305	0.0291	0.00264	0.00364	0.00621	0.0137	0.279	0.199	0.0142
S18	0.00108	0.000667	0.00143	0.12	0.0302	0.0256	0.00194	0.00282	0.00517	0.012	0.294	0.203	0.0152
S19	0.00123	0.000883	0.00178	0.115	0.0268	0.0266	0.00311	0.00425	0.00674	0.0151	0.282	0.194	0.015
S20	0.00116	0.000906	0.00172	0.112	0.0269	0.0267	0.00214	0.00352	0.00634	0.0143	0.21	0.221	0.0168
S21	0.00135	0.001	0.00165	0.116	0.0254	0.0319	0.00229	0.00343	0.00586	0.0118	0.198	0.222	0.0172
S22	0.00084	0.000489	0.00125	0.119	0.0318	0.0281	0.00167	0.00279	0.00507	0.0129	0.278	0.195	0.0129
S23	0.0011	0.000458	0.00117	0.117	0.0275	0.0246	0.00161	0.00258	0.00489	0.00977	0.295	0.207	0.0131
S24	0.000971	0.000402	0.00109	0.109	0.0326	0.0257	0.00161	0.0027	0.00486	0.0104	0.309	0.209	0.0132
S25	0.00149	0.0011	0.00196	0.0949	0.041	0.0301	0.00265	0.00398	0.00647	0.0127	0.243	0.253	0.0137
S26	0.00115	0.000631	0.00144	0.105	0.0316	0.0249	0.00218	0.00344	0.00605	0.0118	0.309	0.207	0.0134
S27	0.0012	0.000803	0.0016	0.0992	0.0263	0.0303	0.00212	0.00331	0.00646	0.0135	0.219	0.227	0.0132
S28	0.00128	0.000822	0.00177	0.107	0.0301	0.0259	0.0023	0.00355	0.00607	0.0125	0.264	0.229	0.0156
S29	0.00152	0.0011	0.00208	0.152	0.0224	0.0274	0.00213	0.00351	0.00573	0.0137	0.269	0.169	0.0135
S30	0.00103	0.000557	0.00109	0.141	0.0227	0.0253	0.00162	0.00276	0.0048	0.0105	0.256	0.19	0.0149
S31	0.00186	0.0016	0.00284	0.157	0.0197	0.0269	0.00271	0.00476	0.00804	0.0204	0.23	0.217	0.0114
S32	0.000826	0.000547	0.00137	0.145	0.0293	0.0274	0.00161	0.0027	0.00479	0.0117	0.24	0.205	0.0154
S33	0.00224	0.00144	0.00266	0.144	0.0216	0.0204	0.00252	0.00391	0.00634	0.0142	0.25	0.223	0.0146
S34	0.000791	0.000265	0.000869	0.146	0.0283	0.0209	0.00128	0.0022	0.00372	0.00741	0.293	0.192	0.0128
S35	0.00116	0.000655	0.00135	0.135	0.0274	0.025	0.00182	0.0032	0.0055	0.0125	0.296	0.164	0.0119
S36	0.0013	0.000725	0.00142	0.106	0.0234	0.0191	0.00195	0.00303	0.00513	0.0123	0.255	0.229	0.0149
S37	0.000788	0.000661	0.0014	0.147	0.0251	0.0304	0.00168	0.00312	0.00545	0.0121	0.225	0.195	0.0158
S38	0.00101	0.000651	0.00125	0.124	0.024	0.021	0.00191	0.00313	0.00543	0.0125	0.278	0.195	0.0124
S39	0.00138	0.000879	0.00206	0.113	0.0405	0.0231	0.0021	0.00351	0.00834	0.0112	0.223	0.241	0.0125
S40	0.00215	0.00149	0.00261	0.141	0.0212	0.0209	0.00282	0.00413	0.00665	0.0125	0.27	0.22	0.0109
S41	0.000995	0.000778	0.00167	0.119	0.0344	0.0253	0.00229	0.00361	0.00742	0.0143	0.217	0.22	0.018
S42	0.0008	0.0006	0.00122	0.11	0.0225	0.0233	0.00173	0.00327	0.00562	0.0114	0.209	0.267	0.02
S43	0.00126	0.000771	0.00159	0.109	0.0337	0.0284	0.00255	0.00379	0.00665	0.0188	0.266	0.214	0.0142
S44	0.00134	0.000833	0.00158	0.143	0.0264	0.0255	0.00214	0.00355	0.00573	0.0123	0.309	0.157	0.0137
S45	0.000996	0.000593	0.00123	0.1	0.0275	0.0248	0.00542	0.00337	0.00474	0.007	0.26	0.205	0.0175
S46	0.00147	0.00139	0.00261	0.122	0.0274	0.0332	0.00255	0.00452	0.00696	0.0143	0.171	0.261	0.0184
S47	0.000925	0.000719	0.00142	0.129	0.0256	0.0239	0.00176	0.00332	0.00603	0.0171	0.25	0.191	0.0121
S48	0.00145	0.00123	0.00243	0.128	0.0308	0.0305	0.00212	0.00408	0.00656	0.0116	0.189	0.244	0.0143
S49	0.00115	0.000741	0.00149	0.148	0.0231	0.0209	0.00197	0.00332	0.00573	0.0129	0.275	0.171	0.0164
S50	0.00116	0.000746	0.00163	0.114	0.0321	0.029	0.00207	0.00332	0.00582	0.0106	0.267	0.226	0.0187
S51	0.00266	0.00209	0.00346	0.129	0.0211	0.0258	0.00645	0.00531	0.00909	0.0184	0.211	0.214	0.0162
S52	0.00189	0.00166	0.00304	0.142	0.0222	0.0278	0.00271	0.00435	0.00778	0.0153	0.218	0.242	0.0135
S53	0.000749	0.000664	0.00164	0.138	0.0375	0.0366	0.0017	0.003	0.00555	0.0129	0.239	0.191	0.0199
S54	0.00161	0.00116	0.00214	0.115	0.0301	0.0251	0.00312	0.00441	0.00748	0.0158	0.328	0.192	0.0122
S55	0.00168	0.00134	0.00241	0.131	0.0217	0.0331	0.00229	0.00452	0.00741	0.0131	0.217	0.228	0.0156
S56	0.00094	0.000549	0.00119	0.125	0.0285	0.0277	0.00168	0.00279	0.00553	0.0117	0.259	0.197	0.016
S57	0.00152	0.00127	0.00208	0.12	0.016	0.0236	0.0034	0.00352	0.00619	0.0137	0.201	0.188	0.0139
S58	0.000973	0.000796	0.00137	0.135	0.0219	0.0269	0.00191	0.00301	0.0054	0.0109	0.223	0.229	0.0176
S59	0.00034	0.000267	0.000804	0.125	0.0246	0.0287	0.00135	0.00261	0.00501	0.0114	0.211	0.215	0.0188
S60	0.000973	0.000732	0.0014	0.115	0.0285	0.0247	0.00338	0.00317	0.00536	0.0123	0.21	0.261	0.0178
S61	0.00123	0.000859	0.00165	0.131	0.0295	0.0235	0.00256	0.00341	0.00594	0.0137	0.293	0.179	0.0139
S62	0.00175	0.00157	0.00279	0.155	0.024	0.0297	0.00304	0.00438	0.0078	0.0238	0.218	0.153	0.0148
S63	0.000978	0.000674	0.00138	0.11	0.0277	0.0261	0.00187	0.00316	0.00541	0.0114	0.223	0.258	0.0179
S64	0.000912	0.000665	0.00142	0.109	0.0334	0.0265	0.00213	0.00364	0.00608	0.0125	0.262	0.186	0.0174
S65	0.000706	0.000671	0.00135	0.111	0.0241	0.0275	0.00171	0.00334	0.00606	0.0142	0.17	0.258	0.0211
S66	0.00135	0.00102	0.00192	0.119	0.0287	0.0283	0.00296	0.00436	0.00699	0.014	0.253	0.231	0.0168
S67	0.00149	0.00119	0.0021	0.121	0.0267	0.0289	0.00329	0.00413	0.0071	0.0151	0.195	0.238	0.0184
S68	0	7.09E-05	0.000445	0.122	0.0302	0.0333	0.00153	0.00258	0.0048	0.012	0.222	0.172	0.0195
S69	0.000868	0.000501	0.00113	0.161	0.0232	0.0199	0.00166	0.00258	0.00464	0.011	0.324	0.142	0.0132
S70	0.00115	0.000921	0.00185	0.142	0.0254	0.0284	0.00194	0.00325	0.00616	0.0139	0.255	0.206	0.0153
S71	0.00139	0.000806	0.00167	0.12	0.0257	0.0265	0.00227	0.00335	0.0053	0.0108	0.299	0.196	0.0156
S72	0.00127	0.00091	0.00187	0.122	0.0366	0.036	0.00227	0.00354	0.00588	0.0138	0.268	0.209	0.0144
S73	0.00111	0.000876	0.00186	0.126	0.0361	0.0275	0.00222	0.00371	0.0066	0.0161	0.264	0.208	0.0136
S74	0.00104	0.000585	0.00148	0.104	0.0512	0.0324	0.00176	0.00296	0.00496	0.0102	0.254	0.225	0.0137
S75	0.000842	0.000716	0.00153	0.159	0.0293	0.0308	0.00205	0.00307	0.00563	0.0143	0.249	0.173	0.0144
S76	0.00123	0.00112	0.00229	0.14	0.029	0.0332	0.0024	0.00374	0.00606	0.0126	0.191	0.243	0.0153
S77	0.000766	0.000603	0.00127	0.142	0.0271	0.0332	0.00203	0.00312	0.0057	0.0129	0.251	0.189	0.0171
S78	0.000821	0.000846	0.00152	0.162	0.0252	0.0262	0.00231	0.00328	0.00577	0.015	0.249	0.177	0.0156
S79	0.00161	0.00142	0.00251	0.165	0.0276	0.0281	0.00262	0.00411	0.00718	0.018	0.246	0.165	0.0178
S80	0.00105	0.001	0.0017	0.156	0.0283	0.0279	0.00236	0.00353	0.00701	0.0149	0.226	0.179	0.0171
S81	0.000572	0.000674	0.00119	0.163	0.0318	0.0284	0.0022	0.00347	0.00594	0.0158	0.297	0.0834	0.0166
S82	0.000922	0.000706	0.0013	0.164	0.0257	0.025	0.0021	0.00308	0.00502	0.0133	0.279	0.174	0.0152
S83	0.000513	0.0004	0.00099	0.155	0.027	0.0272	0.00171	0.00288	0.00528	0.0131	0.251	0.169	0.0148
S84	0.00139	0.00127	0.00219	0.177	0.0238	0.0301	0.00251	0.00398	0.00724	0.0159	0.265	0.0733	0.0177
S85	0.00245	0.00198	0.00378	0.129	0.035	0.0302	0.00356	0.00544	0.00929	0.0156	0.185	0.213	0.0146
S86	0.00209	0.00162	0.0028	0.135	0.0263	0.0293	0.00299	0.0046	0.00762	0.0174	0.249	0.185	0.0108
S87	0.00125	0.000873	0.00167	0.149	0.0284	0.023	0.00227	0.00349	0.00596	0.014	0.311	0.157	0.0125
S88	0.00091	0.000863	0.00156	0.171	0.0243	0.024	0.00216	0.00332	0.00581	0.0146	0.27	0.146	0.0171
S89	0.000639	0.000665	0.00123	0.157	0.0237	0.0258	0.00185	0.00306	0.00555	0.0139	0.244	0.192	0.0156
S90	0.00123	0.000841	0.00165	0.149	0.0283	0.033	0.00218	0.00327	0.00591	0.0133	0.25	0.159	0.0163
S91	0.00162	0.00132	0.00235	0.15	0.0253	0.0272	0.00259	0.00424	0.00697	0.0151	0.228	0.183	0.0156
S92	0.00133	0.00112	0.00202	0.156	0.0217	0.0254	0.00217	0.00368	0.00679	0.0174	0.248	0.183	0.0126
S93	0.000765	0.0007	0.00149	0.132	0.0259	0.0349	0.00208	0.00337	0.0061	0.0139	0.237	0.214	0.0172
S94	0.00125	0.00103	0.00201	0.14	0.0272	0.0283	0.00247	0.00391	0.00684	0.0149	0.258	0.177	0.0147
S95	0.00122	0.00112	0.00216	0.167	0.0244	0.0286	0.00228	0.00359	0.00646	0.0142	0.22	0.178	0.0168
S96	0.00114	0.000839	0.00159	0.131	0.0286	0.0299	0.00248	0.00402	0.00682	0.0147	0.245	0.176	0.0158
S97	0.000931	0.000798	0.00148	0.157	0.0246	0.0282	0.00216	0.00306	0.00543	0.0137	0.252	0.179	0.0166
S98	0.000697	0.000503	0.00106	0.133	0.0266	0.035	0.00189	0.00274	0.00501	0.0125	0.259	0.187	0.0182
S99	0.000922	0.000671	0.0014	0.118	0.0299	0.0317	0.00219	0.00345	0.00606	0.0135	0.273	0.203	0.015
S100	0.00101	0.000685	0.00126	0.136	0.0244	0.0277	0.00184	0.00293	0.00522	0.0118	0.245	0.198	0.0181
S101	0.000764	0.000417	0.00103	0.139	0.0275	0.0227	0.00164	0.00241	0.0181	0.0156	0.302	0.166	0.0122
S102	0.000425	0.0004	0.000958	0.16	0.0271	0.0253	0.00169	0.00273	0.00507	0.0147	0.256	0.16	0.0154
S103	0.00187	0.00168	0.00271	0.187	0.0279	0.0309	0.00291	0.00446	0.00884	0.017	0.266	0.0618	0.0169
S104	0.00194	0.00155	0.00279	0.134	0.0218	0.0293	0.00291	0.00438	0.00756	0.0168	0.216	0.234	0.0144
S105	0.000907	0.000833	0.0017	0.14	0.0248	0.0332	0.00196	0.00337	0.0063	0.0176	0.177	0.207	0.0188
S106	0.000993	0.000913	0.00164	0.135	0.023	0.0303	0.00191	0.00333	0.00588	0.0135	0.18	0.239	0.0183
S107	0.00104	0.000988	0.00187	0.127	0.0307	0.0377	0.00224	0.0038	0.00716	0.0209	0.158	0.237	0.017
S108	0.0021	0.00163	0.00293	0.129	0.0312	0.032	0.00294	0.00441	0.00672	0.0137	0.177	0.227	0.0159
S109	0.00142	0.00144	0.00278	0.139	0.0298	0.0304	0.00267	0.00441	0.00711	0.0135	0.18	0.243	0.0149
S110	0	0	0	0.123	0.0243	0.0251	0.000106	0.00094	0.00573	0.0123	0.202	0.224	0.0199
S111	0.000907	0.000749	0.00138	0.13	0.024	0.0343	0.00192	0.003	0.00557	0.0144	0.176	0.221	0.0187
S112	0.00189	0.00156	0.00295	0.135	0.0315	0.0302	0.0028	0.00439	0.0075	0.0159	0.194	0.222	0.0161
S113	0.0011	0.00108	0.00248	0.136	0.0332	0.0287	0.00215	0.00382	0.00677	0.0151	0.196	0.241	0.0153
S114	0.00135	0.0012	0.00235	0.128	0.0289	0.0327	0.00239	0.00374	0.00732	0.0146	0.168	0.229	0.0179
S115	0.00174	0.0013	0.00242	0.126	0.027	0.029	0.00254	0.004	0.00718	0.0161	0.224	0.201	0.0168
S116	0.000971	0.000869	0.00169	0.128	0.0258	0.034	0.00199	0.0033	0.00642	0.015	0.177	0.207	0.0196
S117	0.00144	0.00118	0.00227	0.121	0.0362	0.031	0.00233	0.00377	0.00659	0.0136	0.209	0.224	0.0155
S118	0.00181	0.00155	0.00279	0.143	0.0229	0.0311	0.00272	0.00451	0.00921	0.015	0.21	0.239	0.0139
S119	0.00137	0.0012	0.00218	0.14	0.0242	0.0329	0.00235	0.00372	0.0062	0.0146	0.185	0.209	0.0196
S120	0.00118	0.00085	0.00161	0.144	0.0254	0.0333	0.00225	0.00327	0.00644	0.0144	0.224	0.163	0.0167
S121	0.000805	0.000671	0.00156	0.144	0.0268	0.027	0.00185	0.00333	0.00595	0.0122	0.221	0.22	0.0168
S122	0.000978	0.00078	0.00178	0.149	0.0266	0.0303	0.00199	0.00351	0.00629	0.0131	0.21	0.184	0.0182
S123	0.00106	0.00103	0.00211	0.141	0.0237	0.0307	0.00239	0.00474	0.00816	0.0158	0.193	0.21	0.0156
S124	0.00113	0.00101	0.00205	0.121	0.0264	0.0329	0.00221	0.00396	0.0071	0.0135	0.184	0.248	0.022
S125	0.000974	0.000708	0.00158	0.112	0.0313	0.0376	0.00201	0.00327	0.00601	0.0139	0.23	0.218	0.0185
S126	0.00083	0.000512	0.00129	0.152	0.0262	0.0269	0.0016	0.00255	0.0051	0.0121	0.265	0.168	0.0161
S127	0.00151	0.00133	0.00224	0.141	0.0262	0.0363	0.00257	0.00402	0.0075	0.017	0.212	0.178	0.0182
S128	0.00176	0.00117	0.00209	0.123	0.0299	0.0297	0.00275	0.00391	0.0067	0.0125	0.222	0.209	0.0175
S129	0.000779	0.0006	0.00127	0.119	0.0278	0.0275	0.00193	0.00288	0.00558	0.0116	0.188	0.242	0.0174
S130	0.0011	0.00102	0.00198	0.121	0.0271	0.0285	0.00232	0.00384	0.00715	0.0153	0.181	0.245	0.0164

	Adipic acid	Adipate	Lysine	Thymine	Lysine	Acetic acid	Ornithine	Methylglut	2-Hydroxy	N-Acetyl-4	Glutamine	Methionin	Pyruvic aci
S1	0.00494	0.0181	0.0111	0.00806	0.00375	0.00356	0.00455	0.00842	0.0349	0.0474	0.00553	0.0132	0.0149
S2	0.00351	0.0141	0.0127	0.00767	0.0072	0.00272	0.00415	0.00759	0.0311	0.0478	0.00582	0.0105	0.0111
S3	0.00378	0.0161	0.0103	0.0112	0.00389	0.00348	0.00471	0.00846	0.0337	0.0399	0.00454	0.0121	0.0149
S4	0.00354	0.0153	0.0123	0.0104	0.00412	0.00333	0.00498	0.00911	0.036	0.0448	0.00521	0.0127	0.00883
S5	0.00443	0.0166	0.00916	0.00949	0.00346	0.0034	0.00456	0.0092	0.0367	0.0357	0.00435	0.0131	0.0158
S6	0.00351	0.0164	0.00875	0.00862	0.00331	0.00296	0.00418	0.00813	0.0359	0.0377	0.00382	0.0108	0.0109
S7	0.00282	0.0155	0.00979	0.0109	0.00328	0.00249	0.00395	0.00761	0.0335	0.0405	0.00505	0.0129	0.0123
S8	0.00304	0.0157	0.0102	0.00877	0.00367	0.00317	0.00484	0.00845	0.0389	0.0453	0.00498	0.0125	0.00886
S9	0.00402	0.0237	0.0131	0.00952	0.00419	0.00285	0.00463	0.00819	0.0371	0.044	0.00705	0.0142	0.00775
S10	0.00301	0.0135	0.0124	0.00965	0.00371	0.00373	0.00502	0.00839	0.0329	0.0464	0.00485	0.0157	0.0185
S11	0.00354	0.0164	0.0124	0.0136	0.00379	0.00306	0.00492	0.00905	0.0351	0.0478	0.0059	0.015	0.0157
S12	0.00259	0.0119	0.0114	0.00681	0.00376	0.00269	0.00472	0.00917	0.0348	0.0467	0.00594	0.0132	0.0146
S13	0.00277	0.00849	0.0114	0.0135	0.00407	0.00427	0.00548	0.0097	0.0313	0.0411	0.00573	0.0157	0.0162
S14	0.00352	0.0265	0.00552	0.00321	0.00169	0.0017	0.00394	0.00815	0.0424	0.0355	0.00312	0.00639	0.00345
S15	0.00402	0.0301	0.00565	0.00419	0.00172	0.00156	0.00285	0.00678	0.0419	0.0358	0.00322	0.00624	0.00478
S16	0.00385	0.0166	0.00912	0.00744	0.0031	0.00279	0.00407	0.0089	0.0414	0.0381	0.00399	0.0103	0.0162
S17	0.00475	0.0205	0.00935	0.00843	0.00383	0.00346	0.00502	0.00893	0.0354	0.0414	0.00556	0.013	0.00966
S18	0.00427	0.0195	0.00876	0.00876	0.00316	0.0035	0.00483	0.00846	0.035	0.0385	0.00444	0.0117	0.0105
S19	0.00575	0.0216	0.00883	0.00542	0.00344	0.00357	0.00484	0.0089	0.0385	0.0418	0.00498	0.0133	0.01
S20	0.0039	0.0133	0.0139	0.00911	0.00404	0.00379	0.00505	0.00906	0.0336	0.0545	0.00634	0.0159	0.0151
S21	0.00366	0.0128	0.0141	0.0194	0.00477	0.00452	0.00521	0.00836	0.0323	0.0445	0.00593	0.019	0.0177
S22	0.00387	0.0209	0.0105	0.00904	0.00355	0.00303	0.00449	0.00781	0.0361	0.0442	0.00502	0.0118	0.00895
S23	0.00349	0.0239	0.0084	0.0201	0.00294	0.00245	0.00331	0.00583	0.0348	0.0391	0.00401	0.0111	0.0142
S24	0.00407	0.0257	0.00864	0.0103	0.00303	0.00254	0.00384	0.00654	0.0336	0.039	0.00465	0.0101	0.00996
S25	0.00382	0.0176	0.0101	0.00991	0.00628	0.00328	0.0046	0.008	0.0304	0.0441	0.00581	0.0127	0.00712
S26	0.00438	0.0253	0.0082	0.0116	0.00302	0.003	0.00427	0.00712	0.0344	0.0399	0.00455	0.011	0.0124
S27	0.00298	0.0169	0.0132	0.032	0.00412	0.00324	0.0042	0.00686	0.0312	0.0535	0.00607	0.0179	0.0155
S28	0.00403	0.0196	0.00995	0.0106	0.00364	0.00321	0.00447	0.00781	0.0342	0.0414	0.00516	0.0119	0.0145
S29	0.00288	0.0139	0.00924	0.00903	0.0029	0.00224	0.00443	0.00968	0.0417	0.0428	0.00478	0.0125	0.00714
S30	0.00364	0.0171	0.00962	0.0212	0.00289	0.00286	0.00455	0.00866	0.041	0.0432	0.00465	0.0148	0.0123
S31	0.00217	0.00668	0.00886	0.00772	0.00248	0.00278	0.00442	0.00897	0.0352	0.0422	0.00349	0.0111	0.0101
S32	0.00361	0.0151	0.0107	0.0106	0.00485	0.00325	0.00419	0.00776	0.0358	0.0447	0.00483	0.012	0.0106
S33	0.00231	0.012	0.00787	0.00468	0.00285	0.00255	0.00344	0.00736	0.0406	0.0449	0.00349	0.0094	0.00875
S34	0.00426	0.0196	0.00851	0.00777	0.00245	0.00249	0.00326	0.00723	0.0429	0.0447	0.00468	0.00922	0.00965
S35	0.00403	0.0223	0.00903	0.0084	0.00282	0.00271	0.00437	0.00853	0.0408	0.0475	0.00483	0.0115	0.00807
S36	0.00397	0.0212	0.00924	0.0046	0.00304	0.00278	0.00351	0.00705	0.0391	0.05	0.00477	0.0119	0.0147
S37	0.00358	0.0116	0.012	0.0125	0.00496	0.0033	0.00446	0.00868	0.0357	0.0427	0.00496	0.0137	0.0125
S38	0.00389	0.0188	0.0094	0.0112	0.00302	0.00306	0.0053	0.01	0.0419	0.0435	0.00526	0.0134	0.0115
S39	0.00309	0.0159	0.01	0.00785	0.00709	0.0029	0.00386	0.00659	0.0309	0.0511	0.00897	0.00901	0.00717
S40	0.00283	0.0123	0.00749	0.00838	0.00294	0.003	0.00408	0.00838	0.0397	0.0408	0.00364	0.0104	0.0101
S41	0.00375	0.0142	0.0119	0.00712	0.00697	0.00295	0.00505	0.00899	0.0313	0.0464	0.00677	0.0119	0.00919
S42	0.00365	0.0122	0.012	0.01	0.0037	0.00369	0.00481	0.00952	0.0331	0.0463	0.00562	0.0142	0.0136
S43	0.00521	0.0212	0.00883	0.00432	0.00356	0.00201	0.00353	0.00691	0.0276	0.0414	0.00622	0.0114	0.00755
S44	0.00351	0.0196	0.00895	0.00728	0.00289	0.00236	0.00423	0.00824	0.0383	0.0408	0.00452	0.0123	0.00703
S45	0.00486	0.0194	0.0114	0.00909	0.00545	0.00399	0.00367	0.00972	0.0393	0.0466	0.00474	0.0122	0.0153
S46	0.00189	0.00548	0.0143	0.0149	0.00892	0.00241	0.00439	0.00851	0.0262	0.0472	0.00682	0.0143	0.00953
S47	0.00363	0.0151	0.0123	0.0101	0.00307	0.00325	0.00493	0.0094	0.0376	0.0523	0.00573	0.0142	0.00889
S48	0.00239	0.00912	0.012	0.0106	0.0103	0.00256	0.00369	0.00723	0.0328	0.0495	0.00536	0.0145	0.00736
S49	0.00359	0.0157	0.00957	0.00784	0.00278	0.00352	0.0056	0.0105	0.0431	0.044	0.00444	0.0139	0.0146
S50	0.00422	0.018	0.0103	0.00769	0.00373	0.00309	0.00392	0.00776	0.0337	0.0372	0.00478	0.0107	0.014
S51	0.00258	0.00812	0.0105	0.00806	0.00451	0.00413	0.00554	0.0108	0.0331	0.0381	0.00462	0.0134	0.0154
S52	0.00162	0.00603	0.00962	0.00886	0.00299	0.00289	0.00427	0.00886	0.0344	0.0426	0.00453	0.0138	0.0093
S53	0.00349	0.0139	0.0139	0.00644	0.00579	0.00306	0.0049	0.00911	0.0351	0.0474	0.00559	0.0129	0.00729
S54	0.00398	0.0208	0.00631	0.00977	0.00242	0.00256	0.0045	0.0086	0.0354	0.0381	0.00459	0.011	0.0068
S55	0.00243	0.00986	0.0113	0.0152	0.00401	0.00292	0.00411	0.00839	0.0338	0.039	0.00461	0.016	0.0138
S56	0.00377	0.0185	0.0114	0.0134	0.00399	0.00319	0.00419	0.00784	0.0366	0.0443	0.00535	0.0136	0.0119
S57	0.00187	0.00658	0.00778	0.00778	0.00319	0.0025	0.00475	0.00899	0.0303	0.0344	0.00389	0.0125	0.0101
S58	0.00337	0.0114	0.0121	0.0135	0.0042	0.00342	0.00486	0.00946	0.0359	0.0428	0.00534	0.0146	0.0135
S59	0.0036	0.0142	0.0123	0.017	0.00368	0.00304	0.00471	0.00923	0.0361	0.0443	0.00533	0.0139	0.0245
S60	0.00356	0.013	0.0114	0.00886	0.00415	0.00338	0.00486	0.00882	0.031	0.0436	0.00545	0.0129	0.0126
S61	0.00361	0.0183	0.0091	0.00705	0.00311	0.00266	0.00449	0.0088	0.036	0.0444	0.00491	0.0108	0.00949
S62	0.00266	0.00668	0.0129	0.0111	0.00493	0.00325	0.00509	0.0105	0.0316	0.048	0.00587	0.0169	0.0118
S63	0.00349	0.015	0.0119	0.00833	0.00378	0.00326	0.00407	0.00733	0.0303	0.0444	0.00526	0.0124	0.0183
S64	0.00454	0.0188	0.0118	0.00969	0.00463	0.0033	0.0048	0.00925	0.0362	0.0507	0.00617	0.013	0.0165
S65	0.00329	0.00918	0.0135	0.0106	0.00375	0.00372	0.00505	0.00976	0.0309	0.051	0.00606	0.0146	0.0254
S66	0.00327	0.0128	0.0106	0.00792	0.00344	0.00327	0.00432	0.00877	0.032	0.0423	0.00458	0.0106	0.0147
S67	0.00284	0.00955	0.0128	0.00683	0.00364	0.00365	0.00536	0.00915	0.0297	0.0496	0.00585	0.0143	0.015
S68	0.00372	0.0154	0.0143	0.0225	0.00396	0.00291	0.00472	0.00887	0.0316	0.0462	0.00634	0.0165	0.0163
S69	0.00416	0.0172	0.00893	0.0143	0.00268	0.00263	0.0047	0.01	0.0434	0.0415	0.00462	0.0117	0.0132
S70	0.00276	0.013	0.00966	0.012	0.00254	0.00295	0.00463	0.00814	0.0358	0.0441	0.00463	0.0139	0.012
S71	0.00455	0.0205	0.00806	0.00423	0.00282	0.00268	0.00367	0.00922	0.0455	0.0395	0.00386	0.00862	0.018
S72	0.00372	0.0166	0.0121	0.00622	0.00539	0.00242	0.00472	0.00884	0.0336	0.0404	0.00539	0.00891	0.00697
S73	0.00342	0.016	0.00967	0.00618	0.00306	0.00242	0.00494	0.00905	0.0327	0.0427	0.00581	0.00913	0.00564
S74	0.00365	0.0209	0.0109	0.00825	0.00744	0.00204	0.00363	0.00606	0.0292	0.0472	0.00654	0.00785	0.00756
S75	0.00355	0.0123	0.0114	0.00842	0.00404	0.00286	0.00491	0.00982	0.0372	0.0447	0.00491	0.0121	0.00707
S76	0.00223	0.00768	0.013	0.0117	0.00586	0.00246	0.00479	0.00915	0.0298	0.045	0.00622	0.0124	0.011
S77	0.00359	0.0134	0.0125	0.0118	0.005	0.00358	0.00542	0.0101	0.035	0.0386	0.00505	0.0148	0.00799
S78	0.00328	0.00905	0.0107	0.0103	0.00321	0.00261	0.00517	0.0106	0.0355	0.0442	0.00552	0.0128	0.00766
S79	0.00258	0.00876	0.0104	0.00928	0.00435	0.00292	0.00565	0.0109	0.0352	0.0417	0.00502	0.013	0.00713
S80	0.00319	0.00925	0.0116	0.0101	0.00369	0.00283	0.00557	0.0107	0.0367	0.0462	0.00682	0.0128	0.009
S81	0.00419	0.0149	0.011	0.0125	0.00438	0.00257	0.00545	0.0111	0.0378	0.0441	0.00578	0.0136	0.00818
S82	0.0039	0.0119	0.00953	0.00835	0.003	0.00255	0.00537	0.011	0.0378	0.0383	0.00482	0.011	0.00678
S83	0.00394	0.0143	0.0121	0.0111	0.0035	0.00274	0.00488	0.00965	0.0391	0.0466	0.00553	0.0126	0.0105
S84	0.00353	0.0111	0.011	0.0218	0.00456	0.00296	0.00477	0.00976	0.0371	0.0395	0.00523	0.0196	0.00687
S85	0.00186	0.00722	0.013	0.00882	0.0135	0.00308	0.00512	0.0084	0.0305	0.0547	0.00627	0.0118	0.00787
S86	0.00235	0.0132	0.0097	0.0129	0.00442	0.00281	0.00459	0.00887	0.0368	0.0432	0.00513	0.0152	0.00631
S87	0.00338	0.0183	0.00851	0.0064	0.00288	0.00272	0.00495	0.00927	0.0396	0.0433	0.00451	0.0104	0.00535
S88	0.00349	0.0118	0.0101	0.00641	0.00333	0.00288	0.00526	0.0108	0.0385	0.0406	0.00496	0.0125	0.00645
S89	0.00328	0.0103	0.0104	0.00877	0.00324	0.00274	0.00502	0.0107	0.0382	0.0419	0.00502	0.0122	0.00811
S90	0.00344	0.0148	0.0123	0.0101	0.0043	0.00322	0.00468	0.00904	0.0389	0.0423	0.00538	0.0133	0.00736
S91	0.00274	0.0111	0.0117	0.0114	0.00407	0.00296	0.00509	0.00974	0.0378	0.0457	0.00579	0.0142	0.00744
S92	0.00296	0.0105	0.0104	0.0125	0.00306	0.00276	0.00532	0.0106	0.0381	0.0421	0.00432	0.0122	0.00805
S93	0.0032	0.0134	0.0126	0.0103	0.00365	0.00316	0.005	0.0095	0.0345	0.043	0.00448	0.0118	0.00855
S94	0.00313	0.0142	0.00948	0.0131	0.00335	0.00301	0.00494	0.00983	0.038	0.0405	0.00544	0.0133	0.00715
S95	0.00273	0.00756	0.0113	0.00871	0.00376	0.00299	0.00502	0.0106	0.0368	0.039	0.00507	0.014	0.0099
S96	0.00334	0.0161	0.012	0.0128	0.00402	0.00308	0.00498	0.00911	0.0348	0.0477	0.00589	0.0145	0.0109
S97	0.00322	0.0119	0.0104	0.0137	0.00353	0.00281	0.00497	0.00973	0.0361	0.0351	0.00449	0.0122	0.0102
S98	0.00342	0.0156	0.0114	0.0141	0.00355	0.00295	0.00459	0.00857	0.0352	0.0381	0.00481	0.0127	0.0147
S99	0.00379	0.02	0.0104	0.00948	0.00329	0.00279	0.00419	0.00792	0.033	0.0424	0.00511	0.0125	0.00788
S100	0.00341	0.0144	0.0109	0.0137	0.00352	0.00301	0.00421	0.00815	0.0342	0.0392	0.00458	0.0128	0.00955
S101	0.00382	0.0205	0.00815	0.00745	0.00475	0.0025	0.00409	0.00822	0.039	0.0364	0.00425	0.0105	0.0066
S102	0.00398	0.013	0.0115	0.00874	0.00422	0.00318	0.0054	0.0104	0.0373	0.0407	0.00535	0.0132	0.00744
S103	0.00303	0.00896	0.0117	0.0139	0.0036	0.00282	0.00578	0.0116	0.0396	0.0466	0.00595	0.0154	0.00682
S104	0.00217	0.00848	0.01	0.0137	0.00355	0.00291	0.00441	0.00872	0.0333	0.0422	0.00466	0.0141	0.0118
S105	0.00319	0.00895	0.0142	0.0124	0.00467	0.0036	0.0049	0.0092	0.0322	0.044	0.00601	0.0163	0.0166
S106	0.00271	0.00812	0.0124	0.0204	0.00409	0.00343	0.00457	0.00877	0.0314	0.042	0.00538	0.0159	0.0169
S107	0.00294	0.00881	0.0138	0.0142	0.00986	0.00289	0.00447	0.0074	0.0241	0.0473	0.0073	0.0142	0.0118
S108	0.00253	0.0087	0.0138	0.0102	0.00904	0.0031	0.005	0.00876	0.0317	0.0511	0.00729	0.0132	0.00927
S109	0.00144	0.00479	0.0109	0.0101	0.00745	0.00186	0.00356	0.0071	0.0282	0.049	0.00669	0.0111	0.00751
S110	0.00324	0.0118	0.00983	0.0344	0.00259	0.00125	0.00268	0.00697	0.0266	0.0369	0.00293	0.0109	0.0196
S111	0.00355	0.01	0.013	0.0313	0.0044	0.00369	0.00523	0.00882	0.0258	0.0411	0.00599	0.0191	0.017
S112	0.00259	0.00878	0.0105	0.0133	0.00804	0.00324	0.00438	0.0075	0.0263	0.046	0.00552	0.0152	0.00818
S113	0.00218	0.00829	0.0101	0.0116	0.00661	0.0017	0.00364	0.00705	0.0259	0.0466	0.00753	0.00986	0.00829
S114	0.00293	0.00862	0.0135	0.0157	0.00964	0.0039	0.0048	0.00833	0.0286	0.0468	0.00612	0.0173	0.00986
S115	0.00264	0.0123	0.0115	0.0115	0.00359	0.00343	0.00481	0.00846	0.0332	0.0518	0.0058	0.0149	0.0145
S116	0.00301	0.0102	0.0147	0.0174	0.00488	0.00436	0.00546	0.00898	0.0314	0.0509	0.00595	0.0187	0.0158
S117	0.00284	0.0113	0.0124	0.0143	0.00963	0.00319	0.00473	0.00848	0.0315	0.0557	0.00603	0.0145	0.00799
S118	0.00203	0.00546	0.00983	0.0101	0.00349	0.00256	0.00416	0.00876	0.0338	0.044	0.00401	0.0127	0.0166
S119	0.00293	0.00795	0.0134	0.00785	0.00412	0.00377	0.0055	0.0106	0.0354	0.0472	0.0056	0.0154	0.0093
S120	0.00349	0.0134	0.0125	0.0158	0.00447	0.00354	0.00514	0.00963	0.0364	0.0414	0.00569	0.0168	0.00994
S121	0.00329	0.0112	0.012	0.00757	0.00552	0.00352	0.00466	0.00867	0.0345	0.0416	0.00476	0.0128	0.00886
S122	0.00349	0.0101	0.0129	0.0111	0.00667	0.00311	0.00486	0.00978	0.0368	0.043	0.00608	0.0139	0.00871
S123	0.00265	0.00785	0.0131	0.0104	0.00626	0.00363	0.00489	0.00926	0.0323	0.0443	0.00536	0.0161	0.0081
S124	0.00216	0.00783	0.0113	0.0118	0.00523	0.00312	0.00438	0.00824	0.0323	0.0432	0.00484	0.0159	0.00809
S125	0.0035	0.016	0.0126	0.00644	0.00524	0.00368	0.00506	0.00867	0.0335	0.0451	0.00541	0.0145	0.00811
S126	0.00312	0.0177	0.0101	0.0152	0.00335	0.00294	0.00421	0.00808	0.0404	0.0435	0.00454	0.0127	0.00857
S127	0.00292	0.0108	0.0138	0.0144	0.00476	0.0043	0.00582	0.0106	0.0369	0.0485	0.00582	0.0168	0.0107
S128	0.00315	0.0141	0.0128	0.00703	0.00431	0.00382	0.00494	0.00865	0.0394	0.0568	0.006	0.0137	0.0176
S129	0.00318	0.0119	0.0131	0.0167	0.00383	0.00383	0.00478	0.008	0.0361	0.0568	0.00605	0.0161	0.0222
S130	0.00221	0.00919	0.013	0.0193	0.00405	0.00349	0.0049	0.0089	0.0338	0.0568	0.00686	0.0181	0.0137

Glutamine	L-Glutamir	Methylami	Citrate	Citrate	Trimethyla	Dimethylgl	Alpha-D-G	Antoate	3-(3-Hydro	3-hydroph	Phenol	L-Histidine	Syringic ac	Tyrosine	Indoxyl sul	Methyl phe	Indoxyl sul	Indoxyl sul	Benzoate	Hippurate	4-Hydroxy	7-Methylx	1-Methylh	Quinolinic	Trigonellin	Hypoxanth
0.00751	0.0014	0.00117	0.0415	0.0276	0.00243	0.00545	0.00136	0.0317	0.000153	0.000723	0.00126	0.00109	0.000236	0.000198	0.00157	0.00143	0.000625	0.000593	0.000192	0.00016	0.000988	0.000613	0.000265	0.000223	0.000862	1.52E-05
0.0075	0.00185	0.00116	0.0547	0.0312	0.00181	0.00522	0.000901	0.0246	0.000197	0.00116	0.00148	0.00114	0.000284	0.0003	0.002	0.00192	0.000858	0.000612	0.000294	0.000171	0.00129	0.000599	0.000114	6.85E-05	0.000125	2.43E-06
0.00749	0.00119	0.00115	0.0476	0.0262	0.00223	0.00551	0.00139	0.0297	0.000193	0.000793	0.0015	0.00133	0.000296	0.000245	0.00168	0.00167	0.000736	0.000714	0.000211	0.000204	0.000938	0.000753	0.000209	0.000181	9.65E-05	2.41E-05
0.00812	0.00128	0.00128	0.0493	0.0286	0.00197	0.00545	0.00227	0.0321	0.000203	0.00114	0.00141	0.00123	0.000226	0.000247	0.00201	0.00165	0.000784	0.000695	0.000269	0.000238	0.00124	0.000498	0.000186	0.000254	0.000121	3.07E-06
0.00928	0.00145	0.00109	0.0426	0.025	0.00179	0.00489	0.00108	0.0321	0.000146	0.000722	0.00124	0.0011	0.00023	0.000214	0.00154	0.00143	0.000641	0.000565	0.000277	0.00017	0.000924	0.000519	0.000239	0.000262	8.48E-05	0
0.00616	0.000558	0.000518	0.0404	0.0223	0.000996	0.00385	0.00104	0.0316	0.000044	0.000813	0.00108	0.00102	0.000222	0.000115	0.00146	0.00125	0.00058	0.0005	0.00019	6.83E-05	0.000594	0.000746	3.97E-05	0	0	0
0.00894	0.00138	0.000867	0.0474	0.0248	0.00075	0.00434	0.001	0.0272	2.28E-05	0.000713	0.00113	0.000686	0.000131	0.000169	0.00154	0.00139	0.000516	0.000343	2.32E-05	0	0.00102	0.000255	3.09E-05	0	5.59E-05	0
0.0084	0.00138	0.00105	0.0489	0.0263	0.00193	0.00493	0.00192	0.0321	0.000101	0.000813	0.00123	0.00131	0.000215	0.000206	0.00196	0.00174	0.000699	0.000516	0.000276	0.000165	0.00102	0.000534	0.000157	0.00019	9.13E-05	0
0.0125	0.00246	0.0015	0.0537	0.0264	0.00278	0.00674	0.00493	0.0303	0.000148	0.00096	0.00137	0.000912	0.000302	0.000311	0.00172	0.00137	0.000639	0.000425	0.000115	0.000104	0.000978	0.00041	0.000106	1.05E-05	0.000458	0
0.0112	0.00145	0.000859	0.0523	0.031	0.00197	0.00548	0.0011	0.0253	8.79E-05	0.00102	0.00141	0.00142	0.000397	0.000277	0.00209	0.00191	0.000814	0.000693	0.000229	0.00017	0.000388	0.00132	0.000252	0.000205	6.58E-05	0
0.0105	0.00197	0.00147	0.054	0.031	0.00173	0.00605	0.00128	0.0308	0.000186	0.00112	0.00163	0.00125	0.000282	0.000306	0.00229	0.00195	0.00084	0.00062	0.000291	0.000241	0.00129	0.0005	0.000359	8.25E-05	0.000112	0
0.011	0.00152	0.000865	0.0611	0.0311	0.00059	0.00528	0.00102	0.0273	6.38E-05	0.000852	0.00137	0.000965	0.000186	0.000237	0.00197	0.00183	0.000747	0.000459	0.000142	0.000027	0.00122	0.000633	0.000136	0	0	0
0.0105	0.00231	0.00194	0.0614	0.0346	0.00154	0.00602	0.000855	0.0264	0.000109	0.000777	0.00146	0.00148	0.000243	0.000217	0.00183	0.00178	0.000813	0.000711	0.000205	0.000145	0.000849	0.00102	9.88E-05	0.000148	0.000176	0
0.00385	0.000655	0.000388	0.0309	0.0175	0.000321	0.00218	0.00327	0.0418	3.36E-05	0.000594	0.000576	0.000591	8.15E-05	8.94E-05	0.00122	0.00112	0.00043	0.000321	0.000111	2.86E-05	0.000527	0.000158	0.000122	0	0.000214	0
0.00322	0.000586	0.000478	0.0225	0.0143	0.000753	0.00185	0.0011	0.0412	5.15E-05	0.000579	0.000536	0.000513	8.99E-05	0.000101	0.00107	0.000929	0.0004	0.000339	0.000119	6.89E-05	0.000501	0.000178	0.000149	1.26E-05	0	0
0.00696	0.00137	0.00112	0.0414	0.0266	0.00125	0.00429	0.000927	0.0374	0.000138	0.000744	0.00115	0.000952	0.000167	0.000171	0.00152	0.00139	0.000612	0.000509	0.000215	0.000116	0.000952	0.000381	0.000209	0.000202	9.01E-05	3.61E-06
0.00774	0.00131	0.00117	0.0456	0.024	0.00201	0.00479	0.00331	0.0362	0.00018	0.000766	0.00111	0.000724	0.000139	0.000165	0.00133	0.00114	0.000414	0.000357	0.00009	0.000087	0.00104	0.000273	0.000179	7.36E-05	0	0
0.00709	0.000923	0.00065	0.0425	0.0222	0.00189	0.00397	0.00241	0.0332	0.000094	0.000774	0.00103	0.000953	0.000334	0.000198	0.00136	0.00116	0.000526	0.000444	0.000136	6.45E-05	0.000392	0.001	0.000174	0.000209	7.99E-05	0
0.00799	0.00108	0.000842	0.0421	0.0255	0.00184	0.00451	0.00249	0.0396	0.000165	0.000722	0.00104	0.000923	0.000199	0.000163	0.00116	0.00116	0.000612	0.000546	0.000188	0.000106	0.000832	0.000491	0.000196	0.000144	9.63E-05	0
0.0101	0.00135	0.00129	0.0634	0.0349	0.00278	0.00743	0.00135	0.026	0.000235	0.00106	0.00178	0.00201	0.000592	0.000307	0.00244	0.0028	0.00126	0.000853	0.000448	0.000324	0.000958	0.00124	0.000272	0.000383	0.000166	2.41E-05
0.0127	0.0023	0.00177	0.0618	0.0356	0.00449	0.00829	0.00221	0.027	0.000125	0.000945	0.00124	0.00145	0.000411	0.000332	0.00184	0.00194	0.000863	0.000849	0.000129	0.000281	0.000432	0.00132	0.000489	0.000488	0.000225	0.000105
0.00767	0.00119	0.00113	0.0493	0.0253	0.00148	0.00454	0.00255	0.0353	0.000181	0.000785	0.00121	0.000941	0.000219	0.000201	0.00151	0.00142	0.000731	0.000603	0.000207	0.000102	0.000995	0.000388	0.000131	0.000198	0	0
0.00651	0.00114	0.00109	0.0404	0.0225	0.00143	0.00422	0.000886	0.0352	0.000138	0.000611	0.000989	0.000737	0.000135	0.000117	0.00116	0.0011	0.000535	0.00047	7.77E-05	7.54E-05	0.000789	0.000249	8.69E-05	0.00019	0	0
0.00584	0.00122	0.00109	0.0352	0.0201	0.00167	0.00398	0.00146	0.0356	0.000144	0.000626	0.00102	0.000774	0.000191	0.000187	0.00119	0.00107	0.000572	0.00049	0.000184	0.000153	0.000918	0.000301	0.000222	9.05E-05	3.58E-05	7.41E-06
0.00628	0.00119	0.00123	0.0462	0.0235	0.00187	0.00516	0.00269	0.0337	0.000142	0.000726	0.0012	0.000898	0.000168	0.000215	0.00145	0.00133	0.000544	0.000367	3.16E-05	0.000021	0.000944	0.000192	0.000154	0	0	0
0.00545	0.000918	0.000781	0.0332	0.0193	0.00176	0.00351	0.00164	0.0359	0.000119	0.000519	0.000897	0.000833	0.000218	0.000152	0.00105	0.00108	0.000515	0.000472	0.000199	0.000176	0.000532	0.000687	0.000152	0.000179	7.51E-05	8.03E-06
0.00843	0.00143	0.00143	0.058	0.032	0.00202	0.00609	0.00116	0.0287	0.000117	0.000767	0.00135	0.00112	0.000165	0.000183	0.00183	0.00196	0.000803	0.000653	0.00012	1.88E-05	0.00106	0.000289	0.000124	0.000171	0	0
0.00662	0.00138	0.00132	0.045	0.025	0.00189	0.00493	0.00137	0.0322	0.000121	0.000575	0.00114	0.000886	0.00016	0.000222	0.00138	0.00128	0.000539	0.000475	9.63E-05	0.000105	0.000963	0.00037	0.000335	6.07E-05	0.000058	6.39E-06
0.00881	0.000697	0.000416	0.0562	0.0308	2.21E-05	0.00387	0.00215	0.0395	0	0.000692	0.00106	0.000676	5.57E-05	0.000111	0.0015	0.00146	0.000638	0.000455	2.19E-05	0	0.000897	0.000189	6.27E-05	0	0	0
0.00924	0.00134	0.000968	0.0443	0.0283	0.00226	0.00443	0.00112	0.0389	0.000148	0.000777	0.000656	0.00155	0.0003	0.00033	0.0018	0.00171	0.000758	0.00072	0.000261	0.000213	8.22E-05	0.000732	0.000538	0.000682	0.000173	5.23E-05
0.00549	0.000157	0.000241	0.0609	0.0334	0.000772	0.00407	0.000962	0.0318	0	0.000565	0.00075	0.000908	0.000161	0.000103	0.00146	0.00155	0.000609	0.000701	0	0	0.000127	0.000864	0.000036	0.000139	0	0
0.00781	0.000965	0.000945	0.0498	0.0304	0.00149	0.00464	0.00108	0.0367	0.000146	0.000806	0.00128	0.00117	0.00019	0.000178	0.00185	0.00171	0.000791	0.000577	0.000243	0.000112	0.00113	0.000439	0.00014	0.000247	0	0
0.00556	0.000472	0.000685	0.05	0.0362	0.000423	0.00356	0.000829	0.0363	9.4E-06	0.000588	0.000912	0.00063	2.53E-05	6.44E-05	0.0014	0.00142	0.000461	0.000366	0	0	0.00101	0.000294	0.000202	0	0.000639	0
0.00507	0.000809	0.000766	0.0338	0.0265	0.00119	0.00285	0.000784	0.0401	0.000157	0.000848	0.00103	0.00106	0.00019	0.0002	0.00163	0.00145	0.000684	0.000567	0.000307	0.000191	0.000993	0.000422	0.000218	0.000304	0.000188	2.66E-05
0.0071	0.00117	0.00116	0.0445	0.0263	0.00164	0.00508	0.00292	0.0408	0.000158	0.000752	0.0012	0.00103	0.000192	0.0002	0.00154	0.00139	0.00066	0.000517	0.000205	0.000106	0.000971	0.000429	0.000219	0.000218	2.06E-05	1.41E-05
0.00844	0.00115	0.00115	0.0477	0.0354	0.00149	0.0049	0.00108	0.0358	0.000166	0.000636	0.00118	0.00109	0.000185	0.000226	0.00157	0.00154	0.000632	0.00049	0.000215	0.00016	0.00099	0.00043	0.000238	0.00047	0.000161	0
0.00899	0.0013	0.00141	0.0587	0.0365	0.00217	0.00603	0.00178	0.0319	0.00023	0.00102	0.00168	0.00146	0.000319	0.000262	0.00205	0.002	0.000995	0.000831	0.000358	0.000287	0.00133	0.000688	0.000283	0.000348	0.00012	2.93E-06
0.00875	0.00156	0.00144	0.0466	0.0292	0.0023	0.00522	0.00118	0.0336	0.000151	0.000659	0.00119	0.00109	0.000227	0.000261	0.00163	0.00154	0.000655	0.000491	0.000256	0.000109	0.00108	0.000405	0.000281	0.000202	4.78E-05	1.98E-05
0.00798	0.00318	0.00238	0.0614	0.0304	0.00166	0.00443	0.00121	0.0283	0.000225	0.000803	0.000816	0.000771	0.00018	0.000239	0.0014	0.00148	0.000619	0.000253	5.96E-05	4.35E-05	0.00102	0.00027	0	0	0	0
0.00477	0.00029	0.000432	0.0483	0.0264	0.000366	0.00347	0.000709	0.0349	1.03E-05	0.000434	0.000821	0.00076	9.83E-05	7.26E-05	0.0012	0.00123	0.000418	0.000373	0	0	0.000592	0.000498	0	0	0	0
0.0102	0.00198	0.00124	0.0712	0.0344	0.00271	0.00646	0.00308	0.0243	0.000265	0.00119	0.00144	0.00128	0.000401	0.000352	0.00201	0.00188	0.000859	0.000515	0.000294	0.000202	0.00122	0.000581	0.000136	8.08E-05	0.000113	1.63E-06
0.00995	0.00149	0.00135	0.0586	0.033	0.00171	0.00557	0.0011	0.0251	0.000206	0.000981	0.00147	0.00136	0.000269	0.000221	0.00201	0.00192	0.000862	0.00071	0.000264	0.000152	0.00122	0.000557	0.000222	0.000209	0	0
0.00824	0.00126	0.00102	0.0503	0.0331	0.00195	0.00473	0.00245	0.0361	0.000156	0.000691	0.0011	0.000729	0.000173	0.000155	0.00121	0.00128	0.000548	0.000384	0.000124	0	0.00105	0.000215	0	0	0	0
0.00887	0.00113	0.000854	0.049	0.0239	0.00119	0.00427	0.00361	0.0362	9.16E-05	0.000782	0.00115	0.000816	0.000189	0.00019	0.00143	0.00118	0.000544	0.000372	7.03E-05	0	0.000883	0.000274	0.000156	0	0	0
0.00569	0.00122	0.00135	0.0609	0.0355	0.00211	0.00577	0.00047	0.0274	0.000235	0.00112	0.00131	0.00128	0.00024	0.000276	0.00235	0.00208	0.000968	0.000711	0.000213	0.000175	0.00121	0.000447	0.000299	0.000395	0.000107	0
0.00838	0.000818	0.000209	0.0797	0.0334	0.000593	0.00496	0.00486	0.0191	5.2E-06	0.00139	0.00169	0.00105	0.000216	0.000199	0.00264	0.00228	0.00101	0.000432	0	0	0.00141	0.000562	0	0	0	0
0.00894	0.00111	0.000874	0.0593	0.0331	0.00221	0.00518	0.00193	0.0293	0.00018	0.000774	0.00138	0.0016	0.000393	0.000304	0.00201	0.00214	0.00093	0.000643	0.000321	0.000206	0.000486	0.00109	0.000373	0.000436	0.000819	2.08E-05
0.00919	0.00111	0.000885	0.073	0.0366	0.000919	0.00541	0.00272	0.0264	0	0.000885	0.00128	0.000845	0.000196	0.000189	0.00189	0.00157	0.000564	0.000346	0	0	0.00116	0.000277	7.5E-06	0	0	0
0.00983	0.000991	0.000638	0.0448	0.0282	0.00169	0.00431	0.00119	0.0377	0.000101	0.00069	0.00119	0.00126	0.00034	0.000209	0.00156	0.0015	0.000638	0.000448	0.000241	0.000167	0.000399	0.00105	0.000211	0.000264	8.19E-05	0
0.00675	0.00135	0.0012	0.0451	0.0242	0.00165	0.00489	0.00115	0.0303	0.000159	0.000806	0.00137	0.00091	0.000209	0.000212	0.00148	0.00136	0.000646	0.000582	0.000187	0.000142	0.00132	0.000399	0.000238	0.000116	0	0
0.00973	0.00161	0.00179	0.0737	0.037	0.00232	0.00731	0.000914	0.0217	0.000079	0.00145	0.00177	0.00129	0.000262	0.000252	0.00262	0.00198	0.00109	0.00079	0.000402	0.00027	0.00168	0.000941	0.000688	0.000385	0.000311	0.000107
0.00791	0	0.000246	0.0671	0.032	0	0.00492	0.001	0.0253	0	0.000741	0.00108	0.00088	3.13E-05	2.63E-05	0.00165	0.00152	0.000671	0.000423	0	0	0.00101	0.00034	0	0	0	0
0.00854	0.00135	0.000984	0.0522	0.0271	0.00111	0.00441	0.00289	0.0294	0.000135	0.00124	0.00137	0.00102	0.000214	0.000193	0.002	0.00197	0.000862	0.000729	0.000267	0.000085	0.00142	0.000363	0.00018	2.18E-05	0	0
0.00433	7.93E-05	0.000207	0.0353	0.0194	0.000311	0.00249	0.00131	0.0363	2.17E-05	0.000541	0.000725	0.000482	2.53E-05	0.000021	0.00114	0.00114	0.0005	0.000523	2.05E-05	0	0.000703	0.000156	0	0.000068	0	0
0.0106	0.00132	0.000972	0.0613	0.0307	0.00064	0.00513	0.00208	0.0283	2.15E-05	0.000909	0.00143	0.000888	0.000172	0.000145	0.00158	0.00151	0.000706	0.000592	4.06E-05	0	0.00129	0.000332	0.00032	1.45E-05	0	0
0.00903	0.00157	0.00144	0.0522	0.0276	0.00208	0.00581	0.00164	0.0345	0.000159	0.000743	0.00121	0.000993	0.00017	0.000192	0.00148	0.00142	0.000661	0.000489	8.66E-05	4.54E-05	0.000978	0.000349	0.000168	0.000136	3.18E-05	0
0.0086	0.000774	0.000696	0.044	0.0256	0.00037	0.00302	0.000794	0.0259	1.34E-05	0.000728	0.000949	0.000642	7.78E-05	0.0001	0.00135	0.0011	0.000545	0.000475	0.000077	0	0.000864	0.000321	0.000185	9.57E-05	0	0
0.01	0.00195	0.00184	0.0558	0.0299	0.00242	0.006	0.000844	0.0288	9.73E-05	0.000959	0.00124	0.000946	0.000153	0.000195	0.00187	0.00146	0.000607	0.000569	6.66E-05	5.15E-05	0.00106	0.000427	0.000161	0.000231	0.000324	3.81E-06
0.0101	0.00221	0.00185	0.0561	0.0395	0.00181	0.00548	0.00158	0.0316	0.0003	0.0013	0.00169	0.00148	0.000332	0.000382	0.00234	0.00206	0.00103	0.000893	0.000416	0.0003	0.00132	0.000593	0.000427	0.000198	0.000165	0
0.008	0.00163	0.00169	0.0595	0.034	0.00226	0.00586	0.00122	0.0243	0.000233	0.000932	0.00159	0.00135	0.000288	0.0003	0.00193	0.00182	0.000791	0.000555	0.000271	0.000218	0.00137	0.000614	0.000323	0.000314	0.000691	3.28E-05
0.00658	0.000932	0.000889	0.0487	0.0287	0.000936	0.00377	0.00228	0.0313	9.44E-05	0.000765	0.00114	0.00106	0.000191	0.000196	0.00162	0.00161	0.000709	0.000479	0.000192	4.07E-05	0.00103	0.000341	0.000195	4.74E-05	0	0
0.0111	0.00118	0.00127	0.0801	0.0377	0.00138	0.00759	0.00618	0.0233	9.86E-05	0.00123	0.0015	0.00113	0.000162	0.000165	0.00249	0.00238	0.000965	0.000908	0.0001	5.15E-06	0.00155	0.000428	0.000409	3.91E-05	0	0
0.00822	0.00158	0.00141	0.0544	0.0299	0.00152	0.00522	0.00105	0.0247	0.00017	0.00114	0.00164	0.00134	0.000266	0.00027	0.00221	0.00189	0.000848	0.000667	0.000303	0.000125	0.00138	0.000548	0.000407	0.000132	6.93E-05	0
0.00811	0.00176	0.00135	0.0537	0.0333	0.00134	0.00463	0.000938	0.0311	0.000147	0.000925	0.00144	0.00111	0.000232	0.000285	0.00203	0.00188	0.000828	0.000661	0.0003	0.000169	0.00121	0.000507	0.000395	0.000059	0.000335	0
0.0101	0.00188	0.0017	0.0694	0.0428	0.00158	0.00573	0.000865	0.0222	0.000152	0.00127	0.00179	0.00143	0.000245	0.000245	0.00256	0.00235	0.000894	0.000847	0.000259	0.000111	0.00158	0.000534	0.000375	0.000374	0.000741	0
0.00568	0.000746	0.00078	0.0496	0.0275	0.000449	0.00312	0.000958	0.0279	5.13E-05	0.000941	0.00111	0.000839	0.000143	0.000117	0.00176	0.00156	0.000623	0.0005	5.17E-05	0	0.00127	0.000359	0.000137	3.24E-05	0.00107	0
0.00987	0.00166	0.00183	0.0692	0.0363	0.00214	0.00714	0.00129	0.0214	0.000226	0.00126	0.00197	0.00175	0.000629	0.000386	0.00261	0.00237	0.00125	0.000737	0.000359	0.000173	0.00161	0.000647	0.000325	0.00039	0.000241	0
0.0147	0.00308	0.00216	0.0736	0.0364	0.00273	0.00782	0.00607	0.0269	0.00025	0.00144	0.00208	0.0015	0.000448	0.000441	0.00256	0.00211	0.000955	0.00065	0.000262	4.13E-05	0.00139	0.000472	0.000217	0	0	0
0.00629	0.00108	0.000858	0.0281	0.0197	0.00167	0.00315	0.00149	0.0389	0.000114	0.000589	0.000569	0.0014	0.000466	0.000205	0.00176	0.00162	0.00065	0.000558	0.000268	0.000243	4.59E-05	0.000173	0.000246	0.00103	0.000343	9.64E-05
0.00808	0.000436	0.000132	0.046	0.0325	0.000757	0.00282	0.00149	0.0298	2.93E-05	0.00111	0.000441	0.00142	0.000207	0.000238	0.00219	0.00177	0.000723	0.000701	0.00012	6.67E-05	0	0.000339	0.000405	0.000507	5.76E-05	0
0.00404	0.000614	0.000752	0.0376	0.0229	0.00048	0.0026	0.000671	0.0408	7.15E-05	0.000818	0.000912	0.00073	8.28E-05	7.87E-05	0.00148	0.00128	0.000574	0.000461	0.000103	3.86E-05	0.000843	0.000323	0.000138	0.000108	0.00104	0
0.00588	0.00109	0.000577	0.0461	0.0239	0.00404	0.00472	0.00182	0.0301	0.00012	0.00106	0.00109	0.000727	0.000177	0.0002	0.00173	0.00142	0.000637	0.000412	0.000125	3.14E-05	0.00104	0.000322	3.46E-05	0	0	0
0.00639	0.000838	0.000362	0.051	0.0252	0.0049	0.00523	0.00473	0.0315	6.97E-05	0.000888	0.0011	0.000801	0.000163	0.000208	0.00168	0.00141	0.000639	0.000302	4.94E-05	0	0.000788	0.000298	0	0	0	0
0.00728	0.00078	0.000569	0.0492	0.025	0.00504	0.00476	0.00113	0.0296	0.000447	0.00106	0.00102	0.000764	0.000175	0.000292	0.00169	0.00139	0.000675	0.000305	0.000146	8.66E-05	0.000898	0.00026	1.63E-05	0	0.000226	0
0.00721	0.000932	0.00091	0.0509	0.0288	0.00703	0.00622	0.00327	0.0336	0.000164	0.000896	0.00143	0.00139	0.000305	0.000223	0.00198	0.00196	0.000964	0.000748	0.000297	0.000169	0.000982	0.000748	0.000128	0.000209	0.000171	0
0.00793	0.00141	0.000823	0.0646	0.0326	0.00533	0.00659	0.00197	0.0223	8.72E-05	0.00107	0.00122	0.000896	0.000197	0.000218	0.00186	0.00172	0.000726	0.000447	8.05E-06	0	0.00107	0.000337	1.16E-05	0	0	0
0.0101	0.00142	0.00121	0.0486	0.0248	0.00607	0.0072	0.00245	0.0314	0.00018	0.00102	0.00135	0.00102	0.000254	0.000238	0.00183	0.00162	0.000794	0.00072	0.000264	0.000215	0.00128	0.000486	0.000186	0.000273	0.000217	1.82E-05
0.00801	0.00105	0.00102	0.0537	0.0332	0.00475	0.00587	0.00245	0.0325	0.000127	0.000955	0.00139	0.00113	0.000155	0.00019	0.00218	0.00202	0.000826	0.000602	0.000248	1.95E-05	0.00127	0.000365	0.000194	0.000199	0	0
0.00804	0.00113	0.000885	0.0545	0.0303	0.00579	0.00646	0.00452	0.0345	1.55E-05	0.0011	0.00142	0.000775	0.000119	0.00017	0.00193	0.0015	0.000641	0.000476	1.87E-05	0	0.00128	0.000266	0.000189	0	0	0
0.00975	0.0018	0.00126	0.0557	0.0339	0.00482	0.00677	0.00389	0.0332	0.000184	0.00111	0.0014	0.000935	0.000274	0.000282	0.00194	0.00171	0.000716	0.000587	0.000177	0.000119	0.00131	0.000607	0.000189	6.42E-05	3.91E-05	0
0.00941	0.00178	0.00135	0.054	0.031	0.00807	0.00701	0.00947	0.0384	0.00018	0.00104	0.00143	0.000941	0.000241	0.000273	0.00193	0.00172	0.000791	0.000652	0.000182	0.000186	0.00122	0.000495	0.000219	0.000131	0	0
0.00694	0.00148	0.00125	0.0447	0.0267	0.00159	0.00478	0.00239	0.0362	0.000161	0.000835	0.00135	0.000953	0.000228	0.000256	0.00164	0.0014	0.0006	0.000471	0.000175	0.000112	0.00118	0.000361	0.000242	7.57E-05	0	1.04E-05
0.00819	0.0017	0.00146	0.0523	0.0305	0.00194	0.00563	0.00217	0.0362	0.000226	0.00099	0.00155	0.00123	0.00029	0.000299	0.00193	0.00191	0.000884	0.000638	0.000282	0.000161	0.00133	0.000442	0.000346	0.000162	0.000239	0
0.0111	0.000731	0.000794	0.0662	0.0419	0.0101	0.00887	0.0077	0.0394	0.000143	0.00096	0.00139	0.00118	0.000187	0.000154	0.00198	0.00172	0.000815	0.000601	5.33E-05	0	0.000871	0.000247	0	0	0	0
0.00704	0.00134	0.00107	0.0716	0.0354	0.0134	0.00988	0.00128	0.0218	0	0.000757	0.00131	0.000888	0.000229	0.000212	0.00193	0.00201	0.00087	0.000324	0	0	0.00124	0.000621	0	0	0.000125	0
0.00768	0.000529	0.00075	0.0589	0.0312	0.00762	0.00756	0.00327	0.0308	0	0.000696	0.000964	0.000815	2.13E-05	0.000106	0.00174	0.00153	0.000561	0.000371	0	0	0.000958	0.000185	0	0	0.000452	0
0.00593	0.000698	0.000745	0.0389	0.0223	0.00527	0.0048	0.00458	0.0361	9.93E-05	0.000771	0.000913	0.000905	0.000148	0.000155	0.00172	0.00154	0.000655	0.000509	0.000229	0.000105	0.00076	0.000318	0.000124	0.000154	0.000335	0
0.00889	0.00122	0.00124	0.0521	0.0326	0.00624	0.00714	0.00293	0.0344	0.000221	0.000979	0.00146	0.00121	0.000264	0.000274	0.00192	0.00165	0.000748	0.000607	0.000302	0.000232	0.00106	0.000466	0.000215	0.00038	0.000547	1.88E-05
0.00806	0.00113	0.00103	0.0502	0.0308	0.00991	0.00643	0.00204	0.0322	0.000169	0.000925	0.00138	0.00114	0.000233	0.000235	0.00193	0.0017	0.000767	0.000595	0.000274	0.000117	0.00113	0.000396	0.000244	0.000135	0	0
0.00865	0.00137	0.00134	0.0538	0.0288	0.00736	0.00798	0.00361	0.0393	0.000173	0.00124	0.00136	0.00123	0.000257	0.000287	0.00221	0.00164	0.000774	0.000639	0.000185	0.000116	0.00108	0.000467	0.000187	0.000108	0.000262	0
0.00923	0.00151	0.00113	0.059	0.0306	0.00831	0.00785	0.00234	0.0318	4.33E-05	0.000918	0.00129	0.000995	0.000173	0.000245	0.00204	0.00189	0.0008	0.00047	0.000174	5.64E-06	0.00116	0.000472	0.000234	0	0.000229	0
0.00663	0.000721	0.0009	0.0537	0.0316	0.00916	0.00684	0.00182	0.0317	0.000152	0.000568	0.00133	0.00116	0.000214	0.000192	0.00159	0.00172	0.000747	0.000537	0.000104	5.25E-05	0.001	0.000466	6.95E-05	0.000196	8.63E-06	0
0.00685	0.000765	0.000795	0.0515	0.0304	0.00805	0.00645	0.00283	0.0276	0.000133	0.0013	0.00131	0.00129	0.000245	0.00021	0.00235	0.0019	0.00085	0.00081	0.000285	0.000106	0.00098	0.00057	0.000151	0.000135	4.67E-05	0
0.00724	0.00113	0.00109	0.0466	0.0276	0.0106	0.00701	0.00677	0.0409	1.16E-05	0.000677	0.000724	0.00056	1.75E-05	4.96E-05	0.00133	0.00102	0.00036	0.000253	0	0	0.000666	0.000124	1.01E-05	0	3.82E-05	0
0.0103	0.00126	0.00122	0.0636	0.0339	0.0121	0.00813	0.00399	0.0294	0.000131	0.001	0.00138	0.00113	0.000183	0.000184	0.00194	0.00175	0.000742	0.000636	0.000112	4.18E-05	0.00119	0.000365	0.000144	6.56E-05	0	0
0.00995	0.00206	0.00158	0.0563	0.0328	0.00984	0.00865	0.00179	0.0289	0.000145	0.000927	0.00143	0.00105	0.000295	0.000324	0.00193	0.00168	0.00076	0.000519	0.000258	9.69E-05	0.001	0.000464	0.000299	6.72E-05	0.000197	9.01E-06
0.00824	0.00134	0.00115	0.05	0.0288	0.0092	0.00707	0.00197	0.0326	0.00013	0.00109	0.00136	0.00104	0.000215	0.000204	0.00208	0.00173	0.000862	0.000777	0.00025	0.000142	0.00116	0.000435	0.000196	0.000161	7.29E-05	0
0.00891	0.0015	0.00129	0.0477	0.0254	0.00794	0.00731	0.00107	0.0317	0.00015	0.000931	0.00128	0.00101	0.000219	0.000197	0.00171	0.00154	0.000703	0.000657	0.000202	7.14E-05	0.001	0.000439	0.000166	0.000135	0	0
0.00784	0.00131	0.00116	0.0433	0.0241	0.00628	0.00654	0.00228	0.0333	0.000159	0.000922	0.00111	0.000861	0.000148	0.000206	0.00161	0.00132	0.00061	0.000506	0.000179	8.96E-05	0.00087	0.000342	0.000156	9.65E-05	7.14E-05	0
0.00871	0.0014	0.00139	0.0557	0.0301	0.0104	0.00848	0.00275	0.0321	0.000255	0.000989	0.00152	0.0012	0.000297	0.000275	0.00176	0.00157	0.000854	0.000702	0.00027	0.000268	0.00113	0.000548	0.000238	0.00026	9.61E-05	4.15E-06
0.0073	0.00128	0.00106	0.0398	0.0241	0.00479	0.00579	0.00223	0.0371	0.000139	0.000645	0.00106	0.000714	0.000175	0.000163	0.00112	0.00108	0.000598	0.000483	0.000161	9.88E-05	0.000838	0.000336	0.000219	5.29E-05	8.22E-05	0
0.00958	0.00173	0.0013	0.0526	0.0314	0.00772	0.00707	0.004	0.0339	0.000193	0.00116	0.00147	0.00128	0.000282	0.000317	0.0022	0.00195	0.000898	0.000698	0.000363	0.000222	0.00116	0.000435	0.000323	0.0001	0.00037	0
0.00931	0.00142	0.00111	0.0659	0.0345	0.0138	0.00838	0.00792	0.0363	7.63E-05	0.00126	0.00139	0.000983	0.000214	0.000298	0.00233	0.00188	0.000896	0.000694	0.000179	0	0.00125	0.000475	0.000266	8.27E-07	0	0
0.00677	0.000384	0.000695	0.0628	0.0343	0.000551	0.00502	0.00064	0.0274	0.000011	0.000915	0.00102	0.000817	5.47E-05	7.26E-05	0.00177	0.00157	0.000566	0.000429	0	0	0.00112	0.000248	1.51E-05	5.38E-05	0	0
0.0123	0.00212	0.00188	0.0778	0.0391	0.00296	0.00796	0.00306	0.0251	0.000284	0.00124	0.0018	0.0014	0.000314	0.000253	0.00228	0.00214	0.00104	0.000802	0.000277	0.000223	0.00177	0.00061	0.000472	0.000233	0.000119	7.59E-06
0.0106	0.00154	0.00144	0.0681	0.0357	0.00204	0.00649	0.00109	0.0254	0.000192	0.00138	0.00167	0.00134	0.000266	0.000313	0.00243	0.00219	0.00106	0.000978	0.000279	0.000264	0.00148	0.000598	0.000278	0.000235	4.82E-05	1.83E-05
0.0102	0.00246	0.00166	0.0789	0.0436	0.00327	0.00808	0.00108	0.0209	0.000238	0.00111	0.00155	0.00131	0.000375	0.00034	0.00216	0.00218	0.000997	0.000602	0.000183	0.000135	0.00132	0.000538	0.000144	0.000173	3.93E-05	7.78E-06
0.0107	0.00209	0.00147	0.0746	0.0395	0.00316	0.0074	0.0016	0.0212	0.000196	0.00139	0.00162	0.00123	0.000339	0.000355	0.00238	0.00194	0.000977	0.000558	0.000202	0.000209	0.00144	0.00065	0.000142	0.00011	0.000159	2.41E-05
0.00711	0.000678	0.000334	0.075	0.0406	0.000931	0.00519	0.00528	0.0362	0	0.00114	0.000702	0.000124	0	0	0.00197	0.0012	4.53E-05	0	0	0	0.000475	0	0	0	0	0
0.00948	0.000699	0.000672	0.0649	0.0398	0	0.00407	0.00647	0.0583	0.000361	0.00157	0.00142	0.000907	0.000132	0.000119	0.00202	0.00131	0.000419	0.000129	0	0	0	0	0	0	0	0
0.00972	0.00188	0.00184	0.0682	0.0425	0.00408	0.00878	0.00133	0.0305	0.000271	0.00118	0.00142	0.00121	0.000268	0.000282	0.00208	0.00188	0.000894	0.000738	0.000113	0.000186	0.00119	0.000464	0.000204	0.000324	0.000238	0
0.00791	0.00133	0.00151	0.0716	0.0376	0.00207	0.00784	0.00413	0.0234	6.09E-05	0.000905	0.00138	0.00111	0.000191	0.000202	0.00207	0.00186	0.000851	0.000533	0.000111	0	0.00128	0.000434	0.000167	0.000151	0	0
0.00856	0.000486	0	0.0767	0.0364	0.000153	0.00371	0.00438	0.0257	5.82E-05	0.001	0.0013	0.000842	0.000175	0.000177	0.00171	0.00182	0.000781	0.000193	0	0	0.00103	0.000155	0	0	0	0
0.0114	0.00184	0.0019	0.0783	0.0393	0.00267	0.00812	0.0028	0.0228	0.000179	0.00133	0.00175	0.00139	0.000269	0.000299	0.00244	0.00215	0.000964	0.000804	0.000217	0.00015	0.00152	0.000454	0.000114	0.000316	0	0
0.00899	0.000888	0.00111	0.0638	0.0348	0.0013	0.00569	0.000835	0.028	7.23E-05	0.001	0.00138	0.00122	0.000196	0.000219	0.00218	0.00191	0.000824	0.000569	0.000131	0	0.00117	0.000497	0.000114	0.000184	0	0
0.0128	0.00164	0.00123	0.0701	0.038	0.00314	0.00774	0.00304	0.0261	0.000241	0.0015	0.00207	0.00199	0.000609	0.00038	0.00284	0.00255	0.0012	0.00101	0.000402	0.000254	0.000927	0.00178	0.000356	0.000492	0.000188	6.13E-05
0.00787	0.00116	0.000817	0.0616	0.032	0.000921	0.00532	0.00236	0.0266	0	0.00105	0.00114	0.000982	0.000174	0.000201	0.0021	0.00199	0.000817	0.000551	0.000118	0	0.0011	0.000388	0.000199	0	0	0
0.00565	0	0.000124	0.0535	0.0325	0.00561	0.00588	0.000814	0.0301	0	0.000695	0.00105	0.001	0.000091	7.23E-05	0.0016	0.00164	0.000723	0.000555	2.67E-06	0	0.000814	0.000367	0	0	0	0
0.0114	0.00167	0.00159	0.068	0.0373	0.0098	0.0102	0.00405	0.0283	0.000207	0.0013	0.00184	0.00152	0.000295	0.00033	0.00255	0.00218	0.00103	0.000875	0.000311	0.000226	0.00157	0.000695	0.000398	0.000317	0.000201	0.000067
0.0118	0.00201	0.00187	0.0617	0.0334	0.0141	0.0107	0.00449	0.0332	0.000253	0.000881	0.00158	0.00125	0.00028	0.000321	0.0019	0.00173	0.000731	0.00065	0.00028	0.000251	0.00118	0.000463	0.000183	0.000268	0.000246	1.56E-05
0.00862	0.00104	0.0011	0.0605	0.0305	0.00743	0.00843	0.0012	0.0275	0.000184	0.00128	0.00163	0.00148	0.000324	0.000257	0.00243	0.00194	0.00091	0.000681	0.000202	1.37E-06	0.00107	0.00061	8.76E-05	8.19E-05	3.82E-07	0
0.0106	0.00216	0.00142	0.0656	0.0339	0.00715	0.00898	0.00287	0.0321	0.000198	0.00136	0.0016	0.00116	0.000354	0.000379	0.00232	0.00187	0.000887	0.000629	0.000188	0.000108	0.00123	0.000496	0.000154	6.72E-05	0.000135	1.32E-05
0.0117	0.00132	0.00124	0.0767	0.0383	0.0073	0.00982	0.00272	0.026	0.000146	0.00107	0.00158	0.00126	0.000196	0.000187	0.00206	0.00187	0.000828	0.000599	0	0	0.00135	0.000358	2.69E-05	0.000192	0	0
0.00983	0.000915	0.000897	0.0699	0.0328	0.00638	0.00828	0.00342	0.0332	6.16E-05	0.000915	0.00118	0.000983	0.000105	8.97E-05	0.00187	0.0016	0.000709	0.000474	0	0	0.00084	0.000111	0	0	0	0
0.00974	0.00138	0.00132	0.0545	0.0282	0.00648	0.0085	0.00262	0.0281	0.000189	0.00126	0.0014	0.00112	0.000221	0.000218	0.00216	0.00171	0.000798	0.000708	0.000189	0.000182	0.00122	0.000498	0.000209	0.000164	0.000266	9.36E-07
0.00852	0.00116	0.00104	0.047	0.029	0.00165	0.00445	0.00398	0.038	0.00016	0.000978	0.00123	0.00109	0.000237	0.00022	0.00191	0.00165	0.000808	0.000665	0.000242	0.00011	0.000929	0.000385	0.000132	0.000222	0.000089	0
0.0102	0.00118	0.00117	0.0577	0.0339	0.00267	0.00613	0.00431	0.0292	0.000107	0.0012	0.00161	0.00151	0.000312	0.000215	0.00248	0.00219	0.000946	0.000798	0.00025	0.000145	0.00113	0.000893	0.000199	0.000182	1.77E-05	9.4E-06
0.00854	0.00146	0.00146	0.0477	0.0322	0.0021	0.00521	0.000865	0.0305	0.000108	0.000886	0.00111	0.0012	0.00019	0.000199	0.00209	0.0018	0.000681	0.000489	0.000179	8.32E-05	0.00118	0.000437	0.000412	0.000243	0.000346	4.03E-05
0.00932	0.00137	0.00114	0.0499	0.0328	0.00248	0.00447	0.00111	0.0306	0.000201	0.001	0.00151	0.00164	0.000354	0.000249	0.00233	0.00229	0.000932	0.000732	0.000352	0.000291	0.000916	0.000968	0.000337	0.000375	0.000277	2.05E-05
0.01	0.00106	0.00102	0.0587	0.0348	0.000849	0.00404	0.000779	0.026	6.34E-05	0.00102	0.00114	0.00137	0.00021	0.000198	0.00249	0.00244	0.000953	0.000562	0.000236	1.31E-05	0.00117	0.000321	0.000312	0	1.01E-06	0

Table 51. NPC1 Dataset encompassing the 130 rows and 55 features peaks values.

2. SVA FEATURES PLOTS FOR THE NPC1 DISEASE DIAGNOSIS

2.1. Plot of NPC1 Disease Dataset

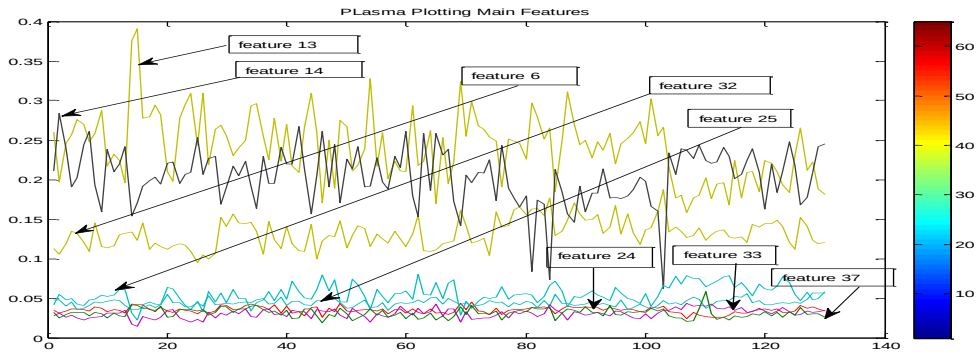

Figure 25. Plot showing the ^{1}H NMR resonance intensities for the most important features found in human plasma i.e. (column number) 13, 14, 6, 32, 25, 24, 33, 37, 28, 27, 34 corresponding to the chemical shift [1.21…1.31], [1.31…1.37], [0.81…0.89], [2.52…2.58], [2.03…2.09], [1.98…2.03], [2.68…2.74], [5.26…5.37], [2.34…2.39], [2.12…2.17], and [2.85…2.88] for NPC1 disease diagnosis. They correspond to the following potential biomarkers (R)-3-hydroxybutyrate; L-fucose; lactate; 3-hydroxyisovalerate, hexacosanoate, L-isoleucine, 2-hydroxyglutarate, Citrate, α-glucose, allantoate, pyruvate, methionine, glutamine, glutamate, and trimethylamine.

Full Boxplot of NPC1 Disease Dataset

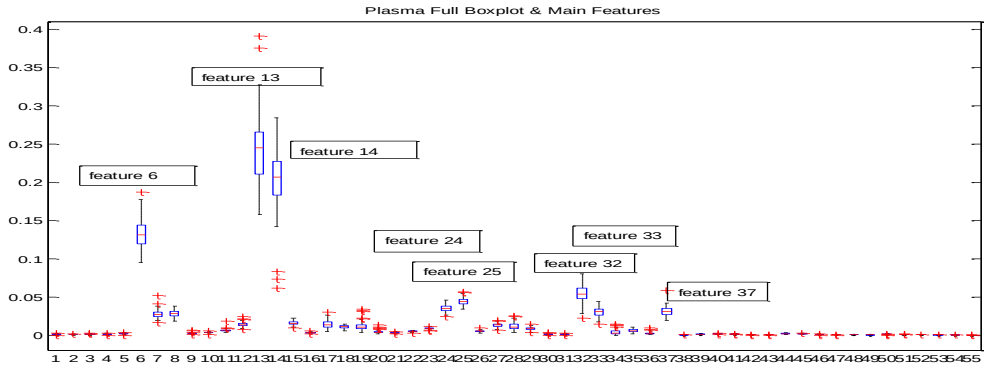

Figure 26. Full Boxplot showing the most important features found in human blood plasma (column number) 13, 14, 6, 32, 25, 24, 33, 37, 28, 27, and 34 corresponding to the chemical shift [1.21…1.31], [1.31…1.37], [0.81…0.89], [2.52…2.58], [2.03…2.09], [1.98…2.03],

[2.68…2.74], [5.26…5.37], [2.34…2.39], [2.12…2.17], and [2.85…2.88] for NPC1 disease diagnosis. They correspond to the following potential biomarkers (R)-3-hydroxybutyrate; L-fucose; lactate; 3-hydroxyisovalerate, hexacosanoate, L-isoleucine, 2-hydroxyglutarate, Citrate, α-glucose, allantoate, pyruvate, methionine, glutamine, glutamate, and trimethylamine.

2.2. Zoomed Plasma Boxplot of NPC1 Disease Dataset

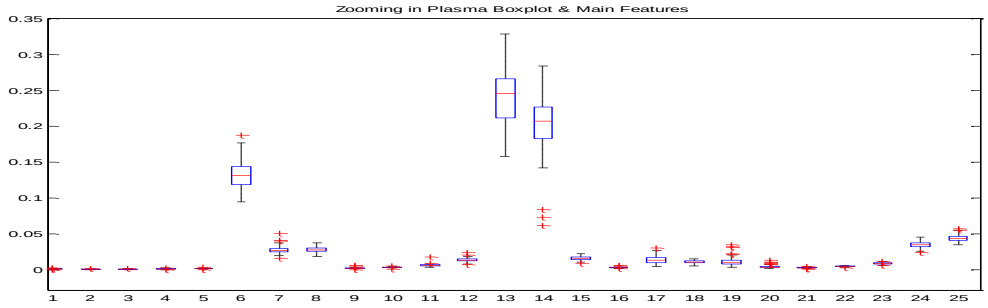

Figure 27. Blood plasma boxplot for important features in the intervals 1-25, showing the significance of features 13, 14 and 6 corresponding to chemical buckets [1.21…1.31], [1.31…1.37], and [0.81…0.89] for NPC1 disease diagnosis. They correspond to the following biomarkers (R)-3-hydroxybutyrate, L-fucose, and lactate.

2.3. Filled Contour Plot of NPC1 Disease Dataset

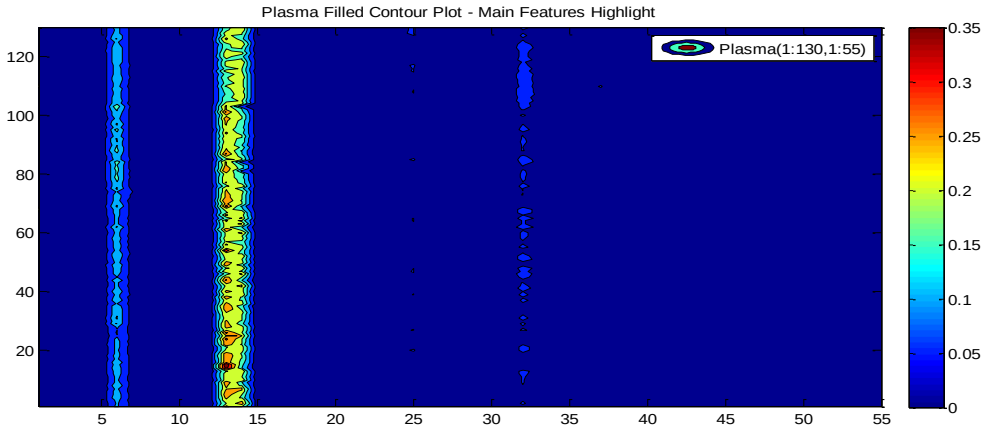

Figure 28. Plasma Filled Contour Plot showing brighter colours for the features 13, 14, 6, 32, 25, 24, 33, 37, 28, 27, 34, corresponding to chemical shift [1.21…1.31], [1.31…1.37], [0.81…0.89], [2.52…2.58], [2.03…2.09], [1.98…2.03], [2.68…2.74], [5.26…5.37], [2.34…2.39], [2.12…2.17], and [2.85…2.88] which are considered as potential biomarkers for NPC1 disease diagnosis. They correspond to the following potential biomarkers (R)-3-hydroxybutyrate; L-fucose; lactate; 3-hydroxyisovalerate, hexacosanoate, L-isoleucine, 2-hydroxyglutarate, Citrate, α-glucose, allantoate, pyruvate, methionine, glutamine, glutamate, and trimethylamine.

2.4. Contour Plot of NPC1 Disease Dataset

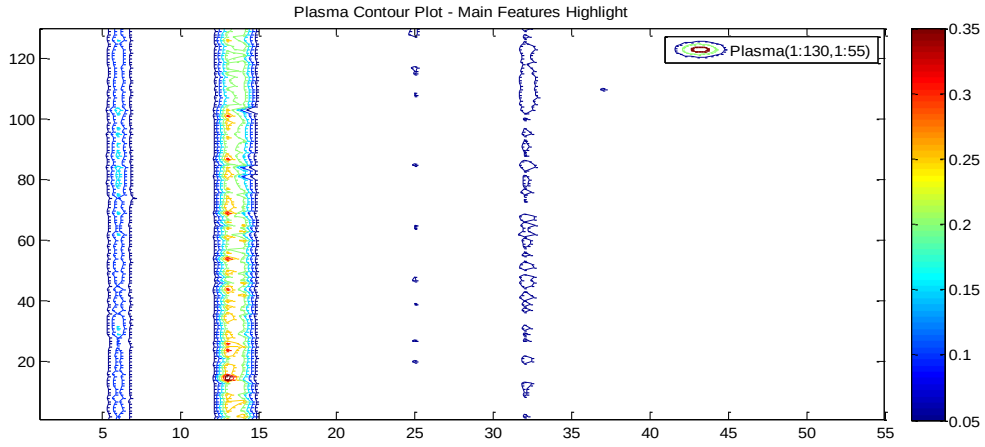

Figure 29. Contour Plot showing brighter colours for the main features (13, 14, 6, 7, 32, 33, 25), corresponding to the chemical shifts [1.21…1.31], [1.31…1.37], [0.81…0.89], [0.89…0.95], [2.52…2.58], [2.68…2.74], and [2.03…2.09], which are potential biomarkers for NPC1 disease diagnosis. They correspond to (R)-3-hydroxybutyrate; L-fucose; lactate; 3-hydroxyisovalerate, hexacosanoate, L-isoleucine, Citrate, etc., considered as main biomarkers, according to the present technique.

3. SVA FEATURES PLOTS FOR THE NPC LIVER DYSFUNCTION DISEASE (NPC LDD) DIAGNOSIS

3.1. Chemical shifts/buckets of biomolecules related to NPC LDD dataset

Chemical Shift	Molecules Name
[0.86 .. 0.92]	(2R)-2amino-4-methylpentanoic acid
[0.92 .. 0.94]	(2R)-2amino-4-methylpentanoic acid
[0.94 .. 0.99]	Leucine-CH_3 - Isoleucine
[1.30 .. 1.36]	Lactate
[1.45 .. 1.51]	Alanine
[1.67 .. 1.70]	Unidentified
[1.70 .. 1.74]	Lysine-C_5-CH_2
[1.74 .. 1.79]	Unidentified
[1.89 .. 1.95]	Acetate - Gamma-Aminobutyric acid
[1.95 .. 1.97]	**2-Hydroxy-3-methylbutyric acid**
[1.97 .. 1.99]	**4-Acetamidobutanoic acid**
[1.99 .. 2.01]	**Methylglutaric acid**
[2.01 .. 2.04]	N-Acetyl-Laspartic acid
[2.09 .. 2.12]	Glutamic acid - N-Acetylaspartylglutamate
[2.12 .. 2.17]	Glutamic acid – Glutamine
[2.17 .. 2.23]	Glutamic acid – Glutamine
[2.23 .. 2.26]	Unidentified
[2.26 .. 2.32]	Gamma-Aminobutyric acid
[2.32 .. 2.37]	Pyruvate-CH_3 and Glutamate-C_3-CH_2
[2.37 .. 2.42]	Glutamate-C_3-CH_2 - Succinate-CH_2's
[2.42 .. 2.47]	Glutamine
[2.47 .. 2.49]	Unidentified
[2.62 .. 2.67]	Methionine - Hypotaurine
[2.67 .. 2.72]	Citrate-CH_2A / CH_2B
[2.72 .. 2.77]	Citrate
[2.79 .. 2.81]	Aspartate
[2.81 .. 2.87]	Aspartate
[2.92 .. 2.97]	6-Methyladenine
[2.97 .. 2.99]	Gamma-Aminobutyric acid
[2.99 .. 3.05]	Lysine - Ornithine
[3.05 .. 3.11]	Ornithine-CS
[3.11 .. 3.15]	Glucuronate- C_2-CH
[3.15 .. 3.17]	9-Methyluric acid
[3.17 .. 3.19]	Unidentified
[3.19 .. 3.21]	Choline
[3.21 .. 3.26]	Phosphocholine - Gamma-phophorylcholine
[3.26 .. 3.31]	Taurine - Myo-Inositol
[3.31 .. 3.35]	Cystine-C_3/C_6-CH_2
[3.39 .. 3.44]	Taurine

[3.44 .. 3.50]	Unidentified
[3.50 .. 3.54]	Glycine
[3.54 .. 3.57]	Myo-Inositol
[3.57 .. 3.62]	Myo-Inositol
[3.62 .. 3.66]	Myo-Inositol
[3.71 .. 3.77]	Glutamic acid – Glutamine
[3.77 .. 3.81]	Glutamic acid – Glutamine
[3.81 .. 3.87]	Unidentified
[3.87 .. 3.93]	Glycerophosphocholine
[3.93 .. 3.95]	Creatine
[4.05 .. 4.10]	Myo-inositol
[4.10 .. 4.16]	Lactate
[4.16 .. 4.21]	Phosphorocholine
[4.21 .. 4.24]	Unidentified
[4.35 .. 4.41]	N-Acetyl-L-aspartic acid
[5.88 .. 5.93]	Uridine
[6.09 .. 6.14]	Inosine acid
[6.49 .. 6.53]	Fumarate
[6.78 .. 6.81]	Tyrosine
[6.81 .. 6.84]	p-Aminobenzoic acid
[6.84 .. 6.86]	**4-Hydroxybenzoic acid**
[6.86 .. 6.88]	p-Hydroxyphenylacetic acid
[7.07 .. 7.12]	Tyrosine
[7.17 .. 7.23]	Tyrosine
[7.30 .. 7.35]	Phenylalanine
[7.47 .. 7.52]	Phenylalanine
[7.65 .. 7.70]	Nicotinate
[7.70 .. 7.72]	Nicotinate – Xanthine
[7.72 .. 7.78]	Hippurate-C_4-CH
[7.85 .. 7.91]	Uridine
[7.99 .. 8.01]	Uridine
[8.23 .. 8.29]	Hypoxanthine
[8.29 .. 8.31]	Inosine
[8.69 .. 8.74]	Nicotinamide adenine dinucleotide
[8.92 .. 8.97]	Nicotinate

Table 52. Chemical buckets and name of the metabolites detected and relative to the NPC liver dysfunction associated with the disease

3.2. XLSTAT Generated Full Statistics Generated for NPC LDD Dataset

Observations/variables table: Workbook = Liver_Original.xlsm / Sheet = Liver.1 / Range = Liver.
Cluster rows
Clustering criterion: Determinant(W)
Stop conditions: Iterations = 500 / Convergence = 0.00001
Number of classes: 2
Center: No
Reduce: No
Initial partition: Random
Repetitions: 10
Seed (random numbers): 4444044
Run again:

Summary statistics:

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
[0.77 .. 0.8	65	0	65	0.000	0.027	0.004	0.004
[0.80 .. 0.8	65	0	65	0.000	0.035	0.005	0.006
[0.82 .. 0.8	65	0	65	0.001	0.040	0.007	0.007
[0.84 .. 0.8	65	0	65	0.002	0.043	0.009	0.008
[0.86 .. 0.9	65	0	65	0.013	0.097	0.038	0.016
[0.92 .. 0.9	65	0	65	0.006	0.034	0.015	0.005
[0.94 .. 0.9	65	0	65	0.025	0.178	0.073	0.031
[0.99 .. 1.0	65	0	65	0.010	0.096	0.035	0.019
[1.05 .. 1.0	65	0	65	0.003	0.038	0.013	0.006
[1.13 .. 1.1	65	0	65	0.001	0.049	0.006	0.008
[1.15 .. 1.1	65	0	65	0.001	0.050	0.006	0.009
[1.17 .. 1.1	65	0	65	0.002	0.025	0.007	0.004
[1.19 .. 1.2	65	0	65	0.004	0.034	0.012	0.006
[1.22 .. 1.2	65	0	65	0.003	0.021	0.008	0.004
[1.24 .. 1.2	65	0	65	0.004	0.026	0.010	0.005
[1.26 .. 1.3	65	0	65	0.011	0.074	0.024	0.012
[1.30 .. 1.3	65	0	65	0.039	0.254	0.117	0.043
[1.45 .. 1.5	65	0	65	0.022	0.181	0.068	0.028
[1.67 .. 1.7	65	0	65	0.004	0.029	0.011	0.005
[1.70 .. 1.7	65	0	65	0.011	0.074	0.029	0.012
[1.74 .. 1.7	65	0	65	0.005	0.047	0.016	0.007
[1.89 .. 1.9	65	0	65	0.009	0.074	0.029	0.013
[1.95 .. 1.9	65	0	65	0.002	0.022	0.007	0.003
[1.97 .. 1.9	65	0	65	0.002	0.020	0.006	0.003
[1.99 .. 2.0	65	0	65	0.003	0.027	0.010	0.005
[2.01 .. 2.0	65	0	65	0.004	0.033	0.012	0.006
[2.09 .. 2.1	65	0	65	0.005	0.099	0.016	0.014
[2.12 .. 2.1	65	0	65	0.014	0.084	0.039	0.016
[2.17 .. 2.2	65	0	65	0.007	0.055	0.022	0.010
[2.23 .. 2.2	65	0	65	0.001	0.061	0.009	0.009
[2.26 .. 2.3	65	0	65	0.003	0.039	0.014	0.007
[2.32 .. 2.3	65	0	65	0.004	0.059	0.023	0.012
[2.37 .. 2.4	65	0	65	0.002	0.031	0.010	0.006
[2.42 .. 2.4	65	0	65	0.001	0.028	0.008	0.005
[2.47 .. 2.4	65	0	65	0.000	0.008	0.002	0.002
[2.51 .. 2.5	65	0	65	0.000	0.008	0.002	0.002
[2.53 .. 2.5	65	0	65	0.003	0.046	0.017	0.010
[2.59 .. 2.6	65	0	65	0.000	0.012	0.001	0.002
[2.62 .. 2.6	65	0	65	0.000	0.023	0.008	0.006
[2.67 .. 2.7	65	0	65	0.000	0.016	0.004	0.004
[2.72 .. 2.7	65	0	65	0.000	0.013	0.002	0.003
[2.79 .. 2.8	65	0	65	0.000	0.012	0.003	0.003
[2.81 .. 2.8	65	0	65	0.000	0.015	0.004	0.004
[2.87 .. 2.9	65	0	65	0.000	0.006	0.001	0.001
[2.90 .. 2.9	65	0	65	0.000	0.005	0.001	0.001
[2.92 .. 2.9	65	0	65	0.000	0.035	0.008	0.007
[2.97 .. 2.9	65	0	65	0.000	0.019	0.006	0.004
[2.99 .. 3.0	65	0	65	0.007	0.063	0.025	0.012
[3.05 .. 3.1	65	0	65	0.000	0.028	0.008	0.007
[3.11 .. 3.1	65	0	65	0.000	0.017	0.004	0.004
[3.15 .. 3.1	65	0	65	0.000	0.008	0.002	0.002
[3.17 .. 3.1	65	0	65	0.000	0.011	0.003	0.003
[3.19 .. 3.2	65	0	65	0.004	0.053	0.016	0.008
[3.21 .. 3.2	65	0	65	0.038	0.202	0.105	0.038
[3.26 .. 3.3	65	0	65	0.035	0.227	0.122	0.046
[3.31 .. 3.3	65	0	65	0.000	0.015	0.006	0.004
[3.35 .. 3.3	65	0	65	0.002	0.051	0.017	0.017
[3.37 .. 3.3	65	0	65	0.004	0.030	0.013	0.006
[3.39 .. 3.4	65	0	65	0.050	0.369	0.174	0.070
[3.44 .. 3.5	65	0	65	0.045	0.355	0.162	0.067
[3.50 .. 3.5	65	0	65	0.013	0.087	0.040	0.016
[3.54 .. 3.5	65	0	65	0.019	0.133	0.064	0.024
[3.57 .. 3.6	65	0	65	0.011	0.197	0.086	0.046
[3.62 .. 3.6	65	0	65	0.008	0.157	0.069	0.039
[3.66 .. 3.6	65	0	65	0.007	0.098	0.045	0.023
[3.69 .. 3.7	65	0	65	0.012	0.119	0.058	0.025
[3.71 .. 3.7	65	0	65	0.045	0.398	0.195	0.077

[3.77 .. 3.8	65	0	65	0.035	0.354	0.156	0.064
[3.81 .. 3.8	65	0	65	0.084	0.432	0.214	0.091
[3.87 .. 3.9	65	0	65	0.040	0.312	0.159	0.061
[3.93 .. 3.9	65	0	65	0.005	0.072	0.037	0.017
[3.95 .. 3.9	65	0	65	0.010	0.135	0.062	0.030
[3.99 .. 4.0	65	0	65	0.008	0.064	0.033	0.013
[4.05 .. 4.1	65	0	65	0.006	0.028	0.014	0.005
[4.10 .. 4.1	65	0	65	0.012	0.063	0.031	0.010
[4.16 .. 4.2	65	0	65	0.003	0.025	0.011	0.004
[4.21 .. 4.2	65	0	65	0.001	0.012	0.005	0.002
[4.24 .. 4.2	65	0	65	0.002	0.015	0.006	0.002
[4.27 .. 4.3	65	0	65	0.003	0.020	0.008	0.003
[4.30 .. 4.3	65	0	65	0.002	0.018	0.008	0.003
[4.35 .. 4.4	65	0	65	0.001	0.016	0.005	0.003
[4.41 .. 4.4	65	0	65	0.001	0.020	0.006	0.003
[4.47 .. 4.5	65	0	65	0.000	0.011	0.002	0.003
[4.55 .. 4.6	65	0	65	0.000	0.019	0.003	0.004
[4.61 .. 4.6	65	0	65	0.000	0.007	0.001	0.001
[4.63 .. 4.6	65	0	65	0.012	0.078	0.034	0.016
[4.89 .. 4.9	65	0	65	0.000	0.073	0.002	0.009
[5.17 .. 5.2	65	0	65	0.000	0.006	0.002	0.001
[5.20 .. 5.2	65	0	65	0.000	0.003	0.000	0.000
[5.22 .. 5.2	65	0	65	0.008	0.101	0.047	0.020
[5.38 .. 5.4	65	0	65	0.000	0.133	0.044	0.035
[5.88 .. 5.9	65	0	65	0.000	0.004	0.001	0.001
[6.09 .. 6.1	65	0	65	0.000	0.003	0.000	0.000
[6.49 .. 6.5	65	0	65	0.000	0.001	0.000	0.000
[6.78 .. 6.8	65	0	65	0.000	0.003	0.001	0.001
[6.81 .. 6.8	65	0	65	0.000	0.002	0.000	0.000
[6.84 .. 6.8	65	0	65	0.000	0.001	0.000	0.000
[6.86 .. 6.8	65	0	65	0.000	0.003	0.000	0.000
[6.88 .. 6.9	65	0	65	0.001	0.076	0.005	0.009
[7.04 .. 7.0	65	0	65	0.000	0.030	0.001	0.004
[7.07 .. 7.1	65	0	65	0.000	0.031	0.004	0.006
[7.12 .. 7.1	65	0	65	0.000	0.004	0.001	0.001
[7.15 .. 7.1	65	0	65	0.000	0.006	0.001	0.001
[7.17 .. 7.2	65	0	65	0.003	0.016	0.007	0.003
[7.23 .. 7.2	65	0	65	0.001	0.075	0.003	0.009
[7.28 .. 7.3	65	0	65	0.000	0.004	0.001	0.001
[7.30 .. 7.3	65	0	65	0.002	0.030	0.007	0.004
[7.35 .. 7.4	65	0	65	0.002	0.185	0.007	0.023
[7.41 .. 7.4	65	0	65	0.002	0.055	0.009	0.010
[7.47 .. 7.5	65	0	65	0.000	0.052	0.002	0.007
[7.52 .. 7.5	65	0	65	0.000	0.081	0.004	0.010
[7.58 .. 7.6	65	0	65	0.001	0.016	0.003	0.002
[7.62 .. 7.6	65	0	65	0.000	0.019	0.001	0.002
[7.65 .. 7.7	65	0	65	0.000	0.027	0.001	0.003
[7.70 .. 7.7	65	0	65	0.000	0.007	0.001	0.001
[7.72 .. 7.7	65	0	65	0.000	0.031	0.002	0.004
[7.78 .. 7.8	65	0	65	0.000	0.007	0.001	0.002
[7.83 .. 7.8	65	0	65	0.000	0.003	0.001	0.001
[7.85 .. 7.9	65	0	65	0.001	0.045	0.006	0.005
[7.91 .. 7.9	65	0	65	0.000	0.039	0.001	0.005
[7.94 .. 7.9	65	0	65	0.001	0.010	0.004	0.002
[7.99 .. 8.0	65	0	65	0.000	0.002	0.001	0.000
[8.01 .. 8.0	65	0	65	0.000	0.002	0.001	0.000
[8.04 .. 8.0	65	0	65	0.000	0.006	0.002	0.001
[8.08 .. 8.1	65	0	65	0.000	0.012	0.001	0.001
[8.12 .. 8.1	65	0	65	0.000	0.014	0.001	0.002
[8.15 .. 8.1	65	0	65	0.000	0.003	0.001	0.001
[8.17 .. 8.2	65	0	65	0.002	0.017	0.008	0.003
[8.23 .. 8.2	65	0	65	0.002	0.019	0.006	0.002
[8.29 .. 8.3	65	0	65	0.000	0.002	0.001	0.000
[8.31 .. 8.3	65	0	65	0.000	0.002	0.001	0.000
[8.33 .. 8.3	65	0	65	0.001	0.012	0.002	0.002
[8.36 .. 8.3	65	0	65	0.000	0.005	0.001	0.001
[8.39 .. 8.4	65	0	65	0.000	0.078	0.003	0.010
[8.44 .. 8.4	65	0	65	0.000	0.007	0.001	0.001
[8.47 .. 8.5	65	0	65	0.001	0.084	0.003	0.011
[8.69 .. 8.7	65	0	65	0.001	0.026	0.003	0.003
[8.74 .. 8.7	65	0	65	0.000	0.003	0.001	0.000
[8.77 .. 8.8	65	0	65	0.000	0.002	0.001	0.000
[8.80 .. 8.8	65	0	65	0.000	0.005	0.001	0.001
[8.86 .. 8.9	65	0	65	0.000	0.003	0.001	0.001
[8.90 .. 8.9	65	0	65	0.000	0.002	0.001	0.000
[8.92 .. 8.9	65	0	65	0.001	0.007	0.002	0.001

Table 53. Statistics Related to the NPC LDD, with the values of the means and standard deviation of the features provided.

3.3. NPC Liver Dysfunction Dataset

Samp+RC	Time	Genotype	[0.77 .. 0.8	[0.80 .. 0.8	[0.82 .. 0.8	[0.84 .. 0.8	[0.86 .. 0.9	[0.92 .. 0.9	[0.94 .. 0.9	[0.99 .. 1.0	[1.05 .. 1.0	[1.13 .. 1.1	[1.15 .. 1.1	[1.17 .. 1.1	[1.19 .. 1.2	[1.22 .. 1.2	[1.24 .. 1.2	[1.26 .. 1.3	[1.30 .. 1.3	[1.45 .. 1.5	[1.67 .. 1.7	[1.70 .. 1.7	[1.74 .. 1.7	[1.89 .. 1.9	[1.95 .. 1.9	[1.97 .. 1.9	[1.99 .. 2.0	[2.01 .. 2.0	[2.09 .. 2.1
1	9	WT/HET	0.0017	0.0015	0.0026	0.0038	0.0195	0.0071	0.0288	0.0111	0.0044	0.0017	0.002	0.0033	0.0061	0.0044	0.0054	0.0129	0.0658	0.0334	0.005	0.0116	0.0071	0.0121	0.0038	0.003	0.0047	0.0059	0.0062
2	9	WT/HET	0.0014	0.002	0.0032	0.005	0.0344	0.0115	0.044	0.0157	0.0057	0.0013	0.0021	0.0034	0.0087	0.0059	0.0075	0.019	0.094	0.0469	0.0055	0.0142	0.0066	0.0163	0.0037	0.0036	0.0058	0.0079	0.0097
3	9	NPC	0.0017	0.0021	0.0032	0.005	0.0303	0.0131	0.0739	0.029	0.0098	0.0019	0.0027	0.0043	0.0091	0.0062	0.007	0.0184	0.1281	0.0735	0.0096	0.0272	0.0133	0.0281	0.0062	0.005	0.0085	0.0112	0.0144
4	9	NPC	0.001	0.0012	0.0024	0.0046	0.0283	0.0085	0.0349	0.0113	0.0033	0.0008	0.0013	0.0022	0.0062	0.0046	0.0067	0.0166	0.0782	0.0334	0.0036	0.0105	0.0049	0.0102	0.003	0.0026	0.0038	0.0063	0.0065
5	11	WT/HET	0.0017	0.0015	0.0025	0.0039	0.0173	0.0073	0.0315	0.0134	0.0055	0.0019	0.002	0.0031	0.006	0.0039	0.0044	0.0114	0.0531	0.0284	0.0059	0.0148	0.008	0.0147	0.0033	0.0028	0.004	0.0046	0.0051
6	11	WT/HET	0.0035	0.0037	0.0047	0.0058	0.0283	0.0109	0.045	0.0194	0.0076	0.0028	0.003	0.0043	0.0079	0.0053	0.0063	0.0163	0.1031	0.0463	0.0077	0.0186	0.0103	0.0207	0.0044	0.004	0.0058	0.007	0.008
7	11	NPC	0.0058	0.0057	0.0073	0.0099	0.0485	0.0198	0.0991	0.0453	0.0186	0.0056	0.0065	0.0094	0.0174	0.0118	0.0128	0.0296	0.1374	0.0946	0.0168	0.0433	0.0243	0.0471	0.0106	0.0094	0.014	0.0171	0.0187
8	11	NPC	0.007	0.0059	0.0076	0.0094	0.0448	0.0186	0.0924	0.0402	0.0168	0.0054	0.0061	0.0083	0.0157	0.0099	0.0104	0.0205	0.1366	0.1018	0.0172	0.0408	0.0252	0.0403	0.0122	0.0114	0.0158	0.019	0.0226
9	11	WT/HET	0.0026	0.0023	0.0033	0.0046	0.0285	0.0106	0.0477	0.019	0.0072	0.0025	0.0028	0.0043	0.0081	0.0056	0.0072	0.0206	0.103	0.0513	0.0074	0.0183	0.0098	0.0191	0.0045	0.0037	0.0062	0.008	0.009
10	9	WT/HET	0.0019	0.0021	0.0028	0.0049	0.0299	0.0098	0.0404	0.015	0.0058	0.002	0.0024	0.0034	0.0075	0.0057	0.0073	0.0186	0.0664	0.0335	0.0058	0.0141	0.0079	0.0148	0.0043	0.0035	0.0054	0.0068	0.0061
11	9	WT/HET	0.0025	0.0024	0.0034	0.005	0.0292	0.0097	0.0368	0.0148	0.0062	0.0025	0.0032	0.0044	0.0084	0.0062	0.0081	0.0232	0.1065	0.0339	0.0066	0.0142	0.0086	0.0147	0.0038	0.0038	0.0061	0.0081	0.0072
12	9	WT/HET	0.0041	0.0043	0.0057	0.0072	0.0405	0.0154	0.0638	0.0245	0.0092	0.003	0.0035	0.0052	0.011	0.0075	0.0084	0.0196	0.1625	0.0796	0.0101	0.0249	0.0141	0.0272	0.0072	0.0066	0.0095	0.0125	0.0138
13	9	WT/HET	0.0054	0.0051	0.0061	0.0067	0.0328	0.0137	0.0638	0.0286	0.0117	0.0043	0.0047	0.0066	0.0113	0.0068	0.007	0.017	0.1597	0.0796	0.0115	0.027	0.0167	0.0331	0.0067	0.0056	0.0088	0.0109	0.0133
14	9	WT/HET	0.0038	0.0035	0.0047	0.0054	0.0272	0.0118	0.0606	0.0252	0.0098	0.0025	0.003	0.0045	0.009	0.0051	0.0056	0.0135	0.1109	0.0679	0.0095	0.0252	0.0139	0.028	0.0059	0.0048	0.0076	0.0096	0.0118
15	11	NPC	0.0002	0.0004	0.0011	0.0024	0.0166	0.0069	0.0301	0.0104	0.003	0.0006	0.0005	0.0015	0.0043	0.0033	0.0048	0.0125	0.0455	0.0216	0.0038	0.0106	0.0052	0.0088	0.0018	0.0019	0.0031	0.0051	0.0087
16	11	NPC	0.0025	0.003	0.0044	0.006	0.0385	0.0152	0.0989	0.0439	0.016	0.003	0.0037	0.0058	0.0116	0.0085	0.0105	0.0264	0.1116	0.0728	0.0115	0.0361	0.0185	0.034	0.0085	0.0069	0.0105	0.0132	0.0141
17	11	NPC	0.0175	0.0348	0.037	0.0427	0.0602	0.0141	0.0678	0.0928	0.0116	0.0357	0.037	0.0152	0.0181	0.0099	0.0087	0.017	0.1142	0.0798	0.0104	0.0269	0.0162	0.0282	0.0094	0.0099	0.0174	0.0236	0.0298
18	11	NPC	0.0062	0.01	0.012	0.0155	0.0457	0.0156	0.0982	0.0578	0.0155	0.0166	0.0174	0.0086	0.0145	0.0098	0.0124	0.0342	0.1324	0.0831	0.0123	0.0361	0.0176	0.0375	0.0089	0.0084	0.0144	0.0186	0.0223
19	11	NPC	0.0061	0.0113	0.0127	0.0163	0.0443	0.0152	0.0986	0.0599	0.0146	0.0101	0.0109	0.0081	0.0126	0.0083	0.0101	0.0257	0.0995	0.0741	0.0094	0.0315	0.0121	0.0604	0.0037	0.0025	0.0047	0.0103	0.0989
21	3	WT/HET	0.0061	0.0053	0.0068	0.0089	0.0445	0.0197	0.0723	0.0333	0.0157	0.0063	0.0063	0.0088	0.0159	0.0109	0.0125	0.028	0.1111	0.0629	0.0178	0.0324	0.0248	0.0315	0.0079	0.0075	0.0118	0.015	0.0141
22	3	WT/HET	0.0025	0.0027	0.0038	0.0053	0.036	0.0112	0.0435	0.0168	0.0066	0.0023	0.0028	0.0043	0.0093	0.0067	0.0092	0.0237	0.0982	0.0479	0.0065	0.016	0.0089	0.0164	0.004	0.0038	0.0059	0.0089	0.0123
23	6	WT/HET	0.006	0.0054	0.0073	0.0086	0.049	0.0164	0.0601	0.0264	0.012	0.0054	0.006	0.0081	0.0168	0.0105	0.0124	0.0311	0.1033	0.0524	0.0111	0.0239	0.0159	0.025	0.0075	0.0072	0.0105	0.0137	0.0159
24	6	WT/HET	0.0055	0.0052	0.0069	0.0089	0.0457	0.0169	0.0689	0.0312	0.0143	0.0061	0.0069	0.0107	0.0207	0.0148	0.017	0.0398	0.1188	0.0589	0.0135	0.033	0.0185	0.0393	0.0079	0.0075	0.0117	0.0141	0.0144
25	9	WT/HET	0.0021	0.0023	0.003	0.0048	0.029	0.0127	0.0657	0.0276	0.0093	0.0017	0.0023	0.0034	0.0082	0.0055	0.0065	0.0186	0.0958	0.063	0.0076	0.0224	0.0107	0.0216	0.0046	0.0036	0.0062	0.0088	0.0109
26	9	WT/HET	0.0058	0.0057	0.0073	0.0086	0.0406	0.0158	0.0661	0.0316	0.0143	0.0057	0.0065	0.009	0.0169	0.0106	0.0112	0.0248	0.1111	0.0562	0.0136	0.034	0.0193	0.0364	0.0074	0.0067	0.0093	0.0113	0.012
27	9	WT/HET	0.0034	0.0035	0.0047	0.0062	0.0339	0.0125	0.0576	0.0249	0.01	0.0033	0.0041	0.0059	0.0124	0.0084	0.0095	0.0238	0.1071	0.0515	0.0099	0.0267	0.0142	0.0256	0.0055	0.0047	0.0072	0.0085	0.0094
28	9	WT/HET	0.003	0.0032	0.0042	0.0054	0.0294	0.0115	0.0534	0.0226	0.0088	0.0027	0.003	0.0049	0.0106	0.0065	0.0071	0.0175	0.0762	0.0508	0.009	0.0233	0.0122	0.0243	0.0049	0.0042	0.0069	0.0093	0.0118
29	9	NPC	0.0056	0.0053	0.0072	0.0109	0.0574	0.0227	0.1217	0.0533	0.0206	0.0058	0.0067	0.0101	0.0209	0.0139	0.0159	0.0384	0.1275	0.0885	0.0206	0.0556	0.0304	0.0529	0.013	0.0114	0.0177	0.0213	0.0237
30	11	NPC	0.0023	0.0026	0.0045	0.0069	0.0443	0.0156	0.0825	0.0326	0.0119	0.0027	0.0033	0.0048	0.0116	0.0088	0.0111	0.0273	0.0962	0.0583	0.0107	0.0306	0.0163	0.0275	0.0087	0.0076	0.0112	0.0149	0.0167
31	11	NPC	0.0008	0.0012	0.0025	0.0049	0.038	0.0159	0.1002	0.0402	0.0133	0.0013	0.0018	0.0037	0.0088	0.0064	0.0077	0.0191	0.0923	0.0645	0.0109	0.0345	0.0156	0.0276	0.0071	0.0061	0.0099	0.0127	0.0146
32	11	WT/HET	0.0016	0.0017	0.0028	0.0047	0.0358	0.0105	0.0462	0.0178	0.0067	0.0023	0.003	0.0044	0.0097	0.0071	0.0102	0.0311	0.1247	0.0713	0.0068	0.0163	0.0091	0.017	0.0033	0.0034	0.0062	0.0091	0.0091
33	3	WT/HET	0.0018	0.0019	0.0031	0.0051	0.0297	0.0125	0.0661	0.0255	0.0089	0.0017	0.0021	0.0037	0.0086	0.0058	0.0065	0.0156	0.0777	0.0573	0.008	0.0247	0.0112	0.0231	0.0049	0.0038	0.0065	0.0094	0.012
34	3	NPC	0.001	0.0012	0.0021	0.0047	0.0287	0.0097	0.047	0.0154	0.0047	0.0009	0.0016	0.0033	0.0078	0.0057	0.0078	0.0171	0.0774	0.0449	0.0051	0.0158	0.0071	0.0181	0.0039	0.0033	0.0053	0.0081	0.0154
35	6	NPC	0.001	0.0015	0.0024	0.0038	0.0204	0.008	0.0361	0.0144	0.0056	0.0019	0.0027	0.0037	0.0085	0.005	0.0058	0.0134	0.0794	0.0458	0.0071	0.0167	0.0098	0.0155	0.0036	0.003	0.0048	0.0065	0.0081
37	6	WT/HET	0.0034	0.0038	0.0049	0.0061	0.033	0.0139	0.0598	0.0245	0.0096	0.0028	0.0034	0.0047	0.0098	0.0067	0.0078	0.022	0.1009	0.0511	0.0097	0.0215	0.0124	0.021	0.0047	0.0041	0.0069	0.0088	0.0095
38	9	NPC	0.004	0.0046	0.0062	0.008	0.0478	0.0194	0.1154	0.0498	0.0185	0.0044	0.005	0.007	0.0131	0.0094	0.0113	0.0289	0.1805	0.0917	0.0161	0.0447	0.0235	0.0395	0.0109	0.0091	0.014	0.0173	0.0202
39	9	NPC	0.0025	0.0025	0.0036	0.0052	0.0271	0.0124	0.0698	0.027	0.0093	0.0015	0.0021	0.0029	0.0068	0.005	0.0057	0.0145	0.0752	0.0505	0.0089	0.0253	0.0125	0.0216	0.0052	0.0044	0.0067	0.0082	0.0091
40	9	WT/HET	0.0048	0.005	0.0063	0.0075	0.0417	0.0142	0.0513	0.0221	0.009	0.0039	0.0044	0.0065	0.0119	0.0089	0.0113	0.0323	0.1355	0.0546	0.0088	0.0194	0.0118	0.0212	0.0052	0.0046	0.0074	0.0094	0.0101
41	9	WT/HET	0.0017	0.0018	0.002	0.0027	0.0131	0.0057	0.0252	0.0236	0.0269	0.0041	0.0032	0.0034	0.0051	0.0036	0.004	0.0106	0.039	0.1805	0.0086	0.021	0.0234	0.0115	0.0029	0.002	0.0029	0.0039	0.0055
42	9	WT/HET	0.0265	0.0348	0.0396	0.0432	0.0919	0.0273	0.0905	0.0818	0.0214	0.0486	0.0501	0.0247	0.0337	0.0201	0.0209	0.0419	0.2157	0.1102	0.019	0.0362	0.0251	0.0439	0.0122	0.0128	0.0205	0.0243	0.0306
43	9	WT/HET	0.0087	0.0136	0.0153	0.0173	0.0441	0.0155	0.069	0.0477	0.0119	0.0087	0.0098	0.0097	0.0153	0.0093	0.0098	0.0238	0.1556	0.0783	0.0098	0.0239	0.0118	0.0482	0.005	0.0048	0.01	0.015	0.0641
44	6	WT/HET	0.0074	0.0112	0.0131	0.0155	0.0421	0.0128	0.0458	0.0343	0.0099	0.0171	0.0182	0.0104	0.0183	0.0108	0.0116	0.0262	0.0932	0.0398	0.0115	0.0285	0.0144	0.0331	0.0051	0.0051	0.008	0.0104	0.0096
45	6	WT/HET	0.0038	0.0042	0.0053	0.006	0.0285	0.0138	0.0648	0.0281	0.0103	0.003	0.0031	0.0046	0.0084	0.0056	0.006	0.0147	0.1001	0.0565	0.0104	0.0246	0.0148	0.0237	0.0055	0.0047	0.0078	0.0098	0.0112
46	3	WT/HET	0.0065	0.0064	0.0079	0.0088	0.0395	0.0173	0.0829	0.0361	0.0145	0.0045	0.0052	0.0072	0.0133	0.0091	0.0103	0.0228	0.1536	0.0813	0.0135	0.0326	0.0191	0.0365	0.0079	0.0071	0.011	0.014	0.0175
47	3	WT/HET	0.0043	0.0049	0.0062	0.0069	0.035	0.0148	0.0777	0.0322	0.0118	0.0029	0.0033	0.0049	0.0099	0.0068	0.0078	0.0167	0.1717	0.0744	0.0105	0.0286	0.0145	0.0335	0.0061	0.0048	0.0082	0.0109	0.0134
48	6	WT/HET	0.0034	0.0035	0.0047	0.0057	0.029	0.0107	0.0477	0.0216	0.009	0.0029	0.0037	0.0055	0.0101	0.0069	0.0086	0.0228	0.0883	0.0424	0.0075	0.0178	0.0099	0.0206	0.0044	0.0043	0.0065	0.0081	0.0094
49	6	WT/HET	0.0038	0.004	0.0052	0.0062	0.0317	0.0132	0.0631	0.0266	0.0101	0.0028	0.0034	0.005	0.0093	0.0055	0.0065	0.017	0.157	0.0643	0.0088	0.0225	0.012	0.0235	0.0049	0.0045	0.0071	0.0093	0.0107
50	9	NPC	0.0033	0.0035	0.0049	0.007	0.0384	0.0194	0.1328	0.0572	0.02	0.0034	0.0039	0.006	0.0116	0.0083	0.0096	0.02	0.1018	0.0882	0.0169	0.0496	0.0252	0.0416	0.0112	0.0094	0.0151	0.0182	0.0214
51	9	NPC	0.0008	0.0012	0.0023	0.0038	0.0234	0.0103	0.0719	0.0305	0.0103	0.001	0.0015	0.0032	0.0069	0.0047	0.0059	0.0155	0.0835	0.0589	0.0081	0.0259	0.0121	0.023	0.005	0.0042	0.0073	0.0096	0.0113
52	11	NPC	0.0011	0.0014	0.0022	0.0034	0.0229	0.01	0.0717	0.0305	0.0098	0.0011	0.0014	0.0031	0.0068	0.005	0.0062	0.014	0.0819	0.0514	0.0074	0.0251	0.0119	0.0211	0.0048	0.004	0.0068	0.0093	0.0116
53	11	NPC	0.0027	0.0033	0.0047	0.0065	0.0361	0.0144	0.0851	0.0385	0.0144	0.0038	0.0046	0.0068	0.0154	0.0148	0.0171	0.0343	0.1214	0.0711	0.0111	0.0318	0.0167	0.0278	0.0077	0.0065	0.0108	0.0136	0.0177
54	11	NPC	0.0027	0.0032	0.0042	0.0061	0.0363	0.0169	0.1154	0.0505	0.0173	0.0027	0.0034	0.0053	0.0105	0.0079	0.0096	0.0218	0.1116	0.0883	0.0149	0.0437	0.0216	0.0363	0.0094	0.0084	0.0133	0.0164	0.0185
55	3	NPC	0.0126	0.0212	0.024	0.0298	0.0523	0.0188	0.1283	0.0956	0.0224	0.0339	0.0356	0.0135	0.0198	0.0117	0.0113	0.0207	0.134	0.1001	0.0158	0.0479	0.0251	0.0451	0.012	0.0109	0.0185	0.0218	0.0272
56	3	WT/HET	0.0168	0.0161	0.0199	0.0214	0.0969	0.034	0.1779	0.089	0.038	0.0134	0.014	0.0195	0.0314	0.0211	0.0252	0.0585	0.2094	0.1398	0.0289	0.0741	0.0473	0.0735	0.0219	0.0198	0.027	0.0326	0.0354
57	6	WT/HET	0.0078	0.0077	0.0102	0.0115	0.0603	0.0224	0.1067	0.0495	0.0205	0.0072	0.0081	0.0109	0.0205	0.0137	0.0165	0.0419	0.1963	0.1001	0.0165	0.0413	0.0241	0.042	0.0108	0.0095	0.0152	0.0195	0.0237
58	9	WT/HET	0.0044	0.0046	0.0062	0.0069	0.0365	0.0147	0.076	0.0338	0.0133	0.0036	0.0041	0.0057	0.011	0.0073	0.0085	0.0209	0.123	0.0672	0.0105	0.0281	0.0152	0.0302	0.0064	0.0056	0.0094	0.0123	0.0154
59	9	WT/HET	0.0036	0.0041	0.0061	0.0087	0.0569	0.0199	0.09	0.0407	0.016	0.0055	0.0067	0.0099	0.0208	0.0169	0.023	0.0655	0.1617	0.0851	0.0131	0.0333	0.0176	0.0323	0.0084	0.0082	0.0144	0.0193	0.0188
60	9	WT/HET	0.0017	0.0021	0.0029	0.0035	0.0174	0.0073	0.0372	0.0169	0.0063	0.0013	0.0018	0.0023	0.005	0.0035	0.0043	0.0108	0.0587	0.0341	0.0048	0.0134	0.0071	0.0116	0.0028	0.0023	0.0038	0.0052	0.0062
61	9	WT/HET	0.0053	0.0053	0.007	0.008	0.0396	0.0161	0.0908	0.0433	0.0173	0.0048	0.006	0.0085	0.0153	0.0102	0.0128	0.0303	0.1546	0.0895	0.0135	0.036	0.0202	0.0359	0.0091	0.0083	0.0131	0.0165	0.0187
62	9	WT/HET	0.0078	0.0078	0.0106	0.0125	0.0634	0.0221	0.1113	0.0531	0.0231	0.0085	0.0094	0.0134	0.0242	0.0176	0.0213	0.0497	0.1784	0.1006	0.0171	0.0431	0.0255	0.047	0.0119	0.0111	0.0172	0.0215	0.0244
63	11	WT/HET	0.0028	0.003	0.004	0.005	0.0278	0.011	0.0626	0.027	0.0099	0.0024	0.0026	0.0042	0.0083	0.0054	0.0062	0.0162	0.0887	0.0466	0.007	0.0206	0.01	0.0194	0.0045	0.0033	0.0058	0.0074	0.0082
64	11	WT/HET	0.0029	0.0032	0.0044	0.0061	0.0325	0.0107	0.0438	0.0172	0.0061	0.0019	0.0022	0.003	0.0076	0.0051	0.0068	0.0175	0.0882	0.039	0.0056	0.0152	0.0074	0.0155	0.0034	0.0035	0.0051	0.0071	0.0084
65	11	WT/HET	0.0064	0.0069	0.0094	0.0117	0.0745	0.0259	0.1181	0.0547	0.022	0.0079	0.0093	0.0134	0.0258	0.0199	0.0262	0.0735	0.2539	0.117	0.0174	0.0441	0.0238	0.0453	0.011	0.0106	0.0184	0.0247	0.0265
66	11	NPC	0.0016	0.0024	0.0035	0.0049	0.0314	0.0148	0.0984	0.042	0.0144	0.0018	0.0029	0.0047	0.0098	0.0068	0.008	0.0181	0.0859	0.0671	0.0106	0.035	0.0163	0.0311	0.0073	0.0062	0.0094	0.0116	0.0124
67	11	NPC	0.0021	0.0029	0.0044	0.0063	0.0423	0.0202	0.132	0.055	0.0186	0.0022	0.0031	0.0055	0.0119	0.0077	0.0098	0.0254	0.2053	0.1088	0.0165	0.0482	0.0227	0.0405	0.009	0.0074	0.0127	0.0171	0.0189

[2.12 .. 2.1	[2.17 .. 2.2	[2.23 .. 2.2	[2.26 .. 2.3	[2.32 .. 2.3	[2.37 .. 2.4	[2.42 .. 2.4	[2.47 .. 2.4	[2.51 .. 2.5	[2.53 .. 2.5	[2.59 .. 2.6	[2.62 .. 2.6	[2.67 .. 2.7	[2.72 .. 2.7	[2.79 .. 2.8	[2.81 .. 2.8	[2.87 .. 2.9	[2.90 .. 2.9	[2.92 .. 2.9	[2.97 .. 2.9	[2.99 .. 3.0	[3.05 .. 3.1	[3.11 .. 3.1	[3.15 .. 3.1	[3.17 .. 3.1	[3.19 .. 3.2	[3.21 .. 3.2	[3.26 .. 3.3	[3.31 .. 3.3	[3.35 .. 3.3
0.0182	0.0095	0.0025	0.0061	0.0092	0.0052	0.0063	0.0017	0.0012	0.0097	0.0002	0.0048	0.0031	0.0018	0.0015	0.0015	0.0006	0.0001	0.0056	0.0029	0.0216	0.0043	0.0013	0.0005	0.0018	0.0115	0.0648	0.0816	0.0028	0.0443
0.0292	0.0147	0.0019	0.0052	0.0164	0.0076	0.0047	0.0005	0.0008	0.0127	0	0.0006	0	0	0	0	0	0	0.003	0.0033	0.0093	0	0	0	0	0.01	0.1154	0.1519	0.0008	0.0412
0.0449	0.0179	0.0049	0.0137	0.026	0.0093	0.0131	0.0033	0.0029	0.0157	0	0.0124	0.0059	0.0005	0.0042	0.0041	0	0	0.0062	0.0067	0.0414	0.0113	0.005	0.0031	0.0042	0.0208	0.1367	0.1562	0.0076	0.0087
0.0217	0.008	0.0006	0.0032	0.0088	0.0023	0.0039	0.0002	0	0.0031	0	0	0	0	0.0004	0	0	0	0	0	0.0204	0	0	0	0	0.0047	0.0598	0.0905	0	0.0375
0.014	0.0076	0.0029	0.0063	0.0057	0.0039	0.0038	0.001	0.001	0.0049	0	0.0028	0.0016	0	0.001	0.0007	0.0003	0	0.0034	0.0029	0.0112	0.0044	0.0029	0.0016	0.0025	0.0108	0.0417	0.0434	0.004	0.0453
0.0283	0.0178	0.0048	0.0093	0.0105	0.0068	0.0078	0.0019	0.0024	0.0145	0.0001	0.0044	0.0025	0.0006	0.0011	0.0016	0.0003	0.0001	0.0066	0.0059	0.0165	0.0051	0.0027	0.0016	0.0026	0.0148	0.0966	0.0881	0.005	0.0065
0.0446	0.0291	0.0115	0.0231	0.0324	0.0165	0.0106	0.004	0.0065	0.0254	0.0037	0.019	0.012	0.009	0.0075	0.0132	0.0034	0.0024	0.0116	0.0101	0.0396	0.0187	0.0135	0.0066	0.0066	0.0186	0.1034	0.1343	0.012	0.0108
0.0534	0.0358	0.0118	0.0246	0.0412	0.0228	0.0142	0.0044	0.0061	0.0392	0.0054	0.0198	0.0126	0.0117	0.0081	0.0133	0.0063	0.0045	0.0299	0.0155	0.0441	0.0238	0.0134	0.0073	0.0094	0.0238	0.1349	0.1723	0.012	0.0477
0.028	0.0138	0.0042	0.0086	0.0115	0.006	0.0086	0.0021	0.0013	0.0108	0	0.0043	0.002	0.0001	0.0014	0.0016	0	0	0.0049	0.0037	0.0159	0.0047	0.0015	0.001	0.0018	0.012	0.0883	0.083	0.0031	0.0056
0.0167	0.0078	0.0029	0.007	0.0057	0.0025	0.0038	0.0008	0.0007	0.005	0	0.002	0.0005	0	0.0012	0.0004	0.0003	0	0.002	0.003	0.0118	0.0032	0.0003	0.0004	0.0013	0.0111	0.0639	0.0437	0.0021	0.0421
0.0218	0.0138	0.0039	0.0074	0.0107	0.0049	0.0046	0.0013	0.0018	0.0145	0.0003	0.0033	0.0014	0.0017	0.0018	0.0012	0.0006	0.0002	0.0081	0.005	0.0122	0.0045	0.0014	0.0009	0.0018	0.0092	0.0836	0.1013	0.0035	0.0422
0.0432	0.0256	0.0051	0.0114	0.0189	0.0095	0.0105	0.0025	0.0029	0.03	0.0002	0.0053	0.0021	0.0024	0.0014	0.0012	0.001	0.0011	0.022	0.0112	0.0267	0.01	0.0048	0.0034	0.0059	0.0212	0.161	0.2242	0.0078	0.0435
0.0396	0.0267	0.0101	0.0145	0.0196	0.0113	0.0122	0.0029	0.004	0.0288	0.0017	0.0095	0.0052	0.0042	0.0033	0.0074	0.0018	0.0023	0.0172	0.0101	0.0283	0.012	0.0072	0.0039	0.005	0.0168	0.1494	0.2127	0.0081	0.0085
0.0352	0.0216	0.0062	0.0117	0.022	0.0097	0.0059	0.0013	0.0029	0.024	0.0004	0.008	0.0048	0.002	0.0024	0.003	0.0006	0.0013	0.0156	0.0082	0.0224	0.0096	0.0064	0.0036	0.0048	0.0155	0.1052	0.1522	0.0055	0.0076
0.0172	0.0071	0.0026	0.0044	0.0102	0.003	0.002	0.0001	0	0.0034	0	0.0002	0	0	0	0	0	0	0	0	0.0068	0.0016	0	0	0	0.004	0.0382	0.0482	0	0.038
0.0351	0.0165	0.0119	0.0156	0.0252	0.0101	0.0058	0.0011	0.0015	0.0095	0	0.0131	0.0068	0.0012	0.0064	0.0046	0	0	0.0004	0.0029	0.0285	0.0116	0.0033	0.0021	0.0021	0.0165	0.085	0.0747	0.0056	0.0069
0.0441	0.0222	0.0089	0.0146	0.0231	0.0121	0.0103	0.0032	0.0043	0.0222	0.0017	0.0125	0.009	0.0059	0.0048	0.0066	0.0024	0.0008	0.0078	0.0071	0.0376	0.0145	0.0072	0.0033	0.0048	0.0216	0.0995	0.1375	0.0097	0.046
0.04+RC:R	0.0232	0.0182	0.0182	0.0328	0.014	0.0094	0.0022	0.0031	0.0192	0.0089	0.0149	0.0084	0.0033	0.0071	0.0072	0.0012	0.0001	0.0067	0.0063	0.0339	0.0144	0.0058	0.0032	0.0036	0.0154	0.0978	0.1232	0.0085	0.0079
0.0836	0.0411	0.0614	0.0283	0.032	0.0155	0.0131	0.004	0.0032	0.0173	0.0062	0.017	0.0084	0.0039	0.0059	0.0069	0.001	0	0.0045	0.0044	0.0286	0.0117	0.0045	0.003	0.0033	0.0163	0.0808	0.0911	0.0072	0.0082
0.0334	0.0257	0.0093	0.0182	0.0194	0.0112	0.0102	0.0034	0.0033	0.022	0.0023	0.0074	0.0051	0.0047	0.0035	0.0046	0.0024	0.0018	0.0138	0.0075	0.0215	0.0179	0.0048	0.0021	0.0031	0.015	0.1404	0.1527	0.0056	0.0414
0.027	0.0151	0.0041	0.0078	0.0179	0.0071	0.0036	0.0003	0.001	0.0122	0	0.0014	0.0003	0	0.0006	0	0	0	0.0053	0.0036	0.0106	0	0	0	0	0.0107	0.1199	0.1283	0.0012	0.043
0.0421	0.0255	0.0074	0.0147	0.0236	0.0131	0.0146	0.0048	0.0045	0.0258	0.0025	0.0078	0.0057	0.0044	0.004	0.0061	0.0029	0.0018	0.0149	0.0095	0.0217	0.0099	0.0054	0.0032	0.005	0.0189	0.1022	0.1185	0.0069	0.0445
0.0395	0.0247	0.0091	0.0145	0.0222	0.0119	0.0094	0.0031	0.0037	0.021	0.0015	0.0085	0.0057	0.0037	0.0043	0.006	0.0013	0.001	0.0097	0.0068	0.0185	0.0083	0.0084	0.0028	0.0039	0.0115	0.1073	0.1393	0.0066	0.0079
0.034	0.0169	0.004	0.0095	0.022	0.0069	0.002	0	0.0009	0.0156	0	0.003	0.0002	0	0.0002	0	0	0	0.0047	0.0037	0.0154	0.002	0	0	0	0.0086	0.1036	0.1311	0.0025	0.003
0.0303	0.0199	0.0086	0.0148	0.0173	0.0101	0.0086	0.0028	0.0032	0.0141	0.0009	0.0072	0.0054	0.0042	0.003	0.0056	0.0011	0.0008	0.0068	0.0057	0.0222	0.008	0.0067	0.0026	0.0035	0.0167	0.0793	0.0907	0.0057	0.0067
0.026	0.0148	0.0073	0.0112	0.0124	0.0076	0.0074	0.0017	0.002	0.0112	0	0.0049	0.0031	0.0015	0.0015	0.0027	0.0001	0	0.0043	0.0041	0.0184	0.0072	0.0053	0.0024	0.0037	0.0161	0.0848	0.097	0.0052	0.0082
0.0261	0.03+RCR(	0.0056	0.0109	0.0219	0.0111	0.0067	0.0009	0.0013	0.0104	0	0.0051	0.0027	0.0004	0.0017	0.0027	0	0	0.0035	0.0031	0.0169	0.0044	0.0021	0.0009	0.0014	0.0094	0.0868	0.1283	0.0036	0.0047
0.0515	0.0296	0.0148	0.0283	0.0467	0.0236	0.0129	0.0042	0.0052	0.0229	0.0038	0.0218	0.0149	0.0096	0.0091	0.0131	0.003	0.0015	0.0094	0.0089	0.0477	0.0235	0.0168	0.0079	0.0102	0.0317	0.1749	0.2269	0.0127	0.0137
0.0404	0.02	0.0054	0.0131	0.0273	0.0129	0.0087	0.002	0.0016	0.0144	0	0.0062	0.0047	0.0008	0.0041	0.0014	0	0	0	0.002	0.0267	0.0083	0	0	0.0007	0.0126	0.0941	0.1254	0.0045	0.0377
0.0374	0.0146	0.0051	0.0121	0.0253	0.009	0.0051	0.0007	0	0.0082	0	0.0067	0.0037	0	0.0036	0	0	0	0	0.0012	0.0275	0.0082	0.001	0.0013	0.0013	0.0149	0.0769	0.1075	0.0039	0.005
0.0244	0.0139	0.0039	0.0083	0.0158	0.0067	0.0041	0.0002	0.0014	0.0106	0	0.0037	0.0014	0	0.002	0	0	0	0.0006	0.0019	0.0127	0.0004	0	0	0	0.0105	0.0943	0.1457	0.0016	0.0037
0.034	0.0145	0.0035	0.0086	0.0254	0.0084	0.0026	0	0.0012	0.0108	0	0.0053	0.0021	0	0.0019	0	0	0	0	0.0024	0.0161	0.0016	0	0	0	0.0097	0.1054	0.1351	0.0025	0.0035
0.0313	0.0157	0.0041	0.007	0.0206	0.0082	0.0046	0.0001	0.001	0.0114	0	0.0031	0.0014	0	0.0022	0	0	0	0	0.002	0.0113	0	0	0	0	0.0053	0.0948	0.1918	0.0025	0.0424
0.0202	0.0075	0.003	0.0063	0.0165	0.0078	0.0063	0.0007	0.0001	0.0031	0.0118	0.0052	0.0019	0	0.0004	0	0	0	0	0.0005	0.0326	0.0022	0	0	0.0003	0.0208	0.0857	0.103	0.0016	0.0046
0.0334	0.0169	0.0046	0.0086	0.0084	0.0029	0.0082	0.0018	0.0013	0.0129	0	0.0015	0	0	0	0	0	0	0.0041	0.0048	0.0141	0.0023	0	0	0	0.0118	0.1033	0.1064	0.0017	0.0037
0.0618	0.0326	0.0105	0.0194	0.0329	0.0151	0.0167	0.005	0.0056	0.0305	0.0021	0.0146	0.0088	0.0035	0.0059	0.0091	0.0022	0.0024	0.0121	0.0111	0.0417	0.018	0.0089	0.0051	0.0056	0.0233	0.1411	0.1367	0.0126	0.0097
0.0278	0.013	0.0044	0.0095	0.0112	0.0035	0.0032	0.0002	0.0003	0.0068	0	0.0036	0.0007	0	0	0	0	0	0	0.0025	0.0176	0.0035	0	0	0	0.0077	0.0647	0.0664	0.0012	0.0032
0.0301	0.0176	0.0059	0.0092	0.0105	0.0043	0.0057	0.0009	0.0015	0.0124	0	0.001	0	0	0.0001	0	0	0	0.0033	0.004	0.0109	0	0	0	0	0.0103	0.0912	0.0913	0.0002	0.0025
0.0227	0.0349	0.0149	0.0111	0.0069	0.0109	0.0208	0.0054	0.0019	0.0064	0.0007	0.0024	0.0016	0.0004	0.0003	0.0007	0	0.0002	0.004	0.0032	0.0077	0.0022	0	0.0005	0.0007	0.0048	0.0381	0.0348	0.0011	0.0041
0.0809	0.0496	0.0102	0.0184	0.0205	0.0123	0.0279	0.008	0.0031	0.0462	0.0013	0.0034	0.0009	0.0011	0.0008	0	0.0011	0.0022	0.0349	0.0187	0.0273	0.0051	0.0006	0.0011	0.0054	0.0246	0.1597	0.1945	0.0022	0.0408
0.067	0.0382	0.0421	0.0181	0.0199	0.0095	0.0112	0.0028	0.0033	0.0263	0.0002	0.0072	0.0028	0.0002	0.0014	0.0007	0	0	0.0117	0.0087	0.0291	0.0055	0.0018	0.0017	0.0026	0.024	0.1026	0.085	0.0057	0.0066
0.0187	0.0118	0.0045	0.0067	0.0039	0.0044	0.0051	0.0006	0.0006	0.0036	0	0.002	0	0	0	0	0	0	0	0.0009	0.0308	0.0025	0.0023	0	0.0004	0.0109	0.0504	0.0633	0.0014	0.0505
0.0334	0.0217	0.007	0.0105	0.0186	0.0069	0.0033	0.0004	0.002	0.0198	0	0.0053	0.0025	0	0.0015	0.0009	0	0.0004	0.011	0.0074	0.0216	0.01	0.0038	0.0025	0.0038	0.0146	0.0921	0.1002	0.0057	0.007
0.0475	0.0308	0.0096	0.0168	0.0293	0.0121	0.0082	0.0024	0.0035	0.0276	0.0008	0.0086	0.0047	0.0015	0.0029	0.0041	0.0012	0.0012	0.0153	0.0107	0.0277	0.0107	0.0054	0.003	0.004	0.0161	0.1432	0.1646	0.008	0.0073
0.0421	0.0251	0.007	0.0108	0.0229	0.0072	0.0037	0.0002	0.0028	0.0237	0	0.0044	0.0018	0	0.0011	0	0	0	0.0087	0.0079	0.0245	0.0051	0	0.0007	0.0014	0.015	0.1542	0.1649	0.006	0.0053
0.0287	0.0162	0.0075	0.0082	0.0145	0.0052	0.0045	0.0011	0.0017	0.0145	0	0.0033	0.0011	0.0002	0.002	0.0019	0	0.0003	0.0081	0.0047	0.0138	0.0038	0.0007	0.0007	0.0014	0.0112	0.0759	0.078	0.0028	0.0055
0.0345	0.0222	0.012	0.0086	0.0175	0.0059	0.0039	0.0008	0.002	0.019	0	0.0033	0.0013	0	0.0003	0	0	0	0.0083	0.0068	0.0189	0.005	0.0011	0.001	0.0018	0.0146	0.1285	0.1376	0.0072	0.0077
0.0518	0.0235	0.0112	0.0231	0.0409	0.016	0.0099	0.0031	0.0036	0.0191	0.0016	0.0181	0.0116	0.0039	0.0082	0.0094	0.0021	0.001	0.0086	0.0085	0.0456	0.0204	0.0099	0.0063	0.0066	0.0195	0.0982	0.1102	0.0114	0.0101
0.0257	0.011	0.0043	0.01	0.0207	0.0097	0.0053	0.0003	0.0007	0.006	0	0.0139	0.003	0	0.0026	0.0049	0	0	0.001	0.0017	0.0257	0.0062	0.0002	0.0004	0.0012	0.0098	0.0801	0.1014	0.0038	0.0056
0.0308	0.0134	0.0059	0.009	0.0245	0.0081	0.0034	0.0004	0.0006	0.0094	0	0.0066	0.0033	0	0.0027	0.0005	0	0	0	0.002	0.0197	0.0055	0	0.0005	0.0002	0.0093	0.0734	0.0812	0.0036	0.005
0.043	0.0237	0.0102	0.0147	0.0326	0.0145	0.0097	0.0026	0.0031	0.0177	0.0012	0.012	0.0079	0.0023	0.0054	0.0053	0.0006	0	0.0043	0.0052	0.0361	0.0115	0.0051	0.0035	0.004	0.0156	0.0933	0.1191	0.0086	0.0079
0.0464	0.0227	0.0099	0.0191	0.035	0.0141	0.0085	0.0021	0.0032	0.0189	0.0009	0.0152	0.0096	0.0029	0.0071	0.0073	0.0009	0.0008	0.008	0.0073	0.0392	0.0163	0.0072	0.0045	0.0048	0.0187	0.0951	0.1013	0.0086	0.0086
0.0559	0.0249	0.0102	0.0222	0.0321	0.0148	0.0153	0.0046	0.0033	0.0177	0.0021	0.0163	0.0106	0.0031	0.0071	0.0078	0.0015	0.0002	0.0051	0.0061	0.0442	0.0201	0.0093	0.0058	0.0062	0.0246	0.0948	0.0993	0.0111	0.0095
0.0754	0.0545	0.0197	0.0388	0.0585	0.0311	0.0166	0.0057	0.008	0.0437	0.0057	0.0233	0.0155	0.0131	0.0122	0.0151	0.0062	0.0042	0.026	0.0193	0.0626	0.0281	0.0131	0.0071	0.0109	0.0527	0.1746	0.2072	0.0151	0.0457
0.066	0.0398	0.0113	0.0224	0.042	0.0171	0.0141	0.0043	0.0059	0.039	0.0019	0.0124	0.0073	0.0042	0.0062	0.0087	0.0024	0.0021	0.0222	0.0148	0.0387	0.0153	0.0086	0.0052	0.0075	0.0264	0.2016	0.2053	0.0123	0.0104
0.0454	0.024	0.0058	0.0134	0.0288	0.0106	0.0083	0.0019	0.003	0.0224	0	0.0068	0.003	0	0.0027	0.0027	0.0004	0.0003	0.0113	0.008	0.0246	0.0085	0.004	0.0025	0.0037	0.0214	0.1359	0.1358	0.007	0.0069
0.0462	0.0276	0.0087	0.0179	0.0299	0.0125	0.0058	0.0011	0.0025	0.0196	0	0.0071	0.0034	0.0015	0.0055	0.0055	0	0	0.0071	0.006	0.0229	0.0054	0	0.0001	0.0006	0.0194	0.0983	0.1124	0.0047	0.0043
0.0192	0.0122	0.0028	0.0057	0.0109	0.004	0.0006	0	0.0006	0.0094	0	0.001	0	0	0	0	0	0	0.001	0.0019	0.008	0	0	0	0	0.0078	0.0601	0.0689	0.0007	0.0022
0.0431	0.0252	0.0095	0.0204	0.0391	0.0168	0.0069	0.0026	0.0046	0.0194	0.0014	0.0124	0.0083	0.0044	0.0062	0.0078	0.0022	0.0014	0.0084	0.0085	0.0354	0.0152	0.0093	0.0053	0.0065	0.0306	0.1264	0.1082	0.0104	0.0096
0.0521	0.0319	0.0131	0.0257	0.0431	0.023	0.0167	0.0041	0.0051	0.0211	0.0028	0.0165	0.0111	0.0079	0.0088	0.0141	0.0028	0.0018	0.0099	0.0091	0.0402	0.0178	0.0084	0.005	0.0064	0.0223	0.141	0.1462	0.011	0.009
0.0275	0.0137	0.0029	0.0066	0.0116	0.0045	0.0038	0.0005	0.0006	0.0106	0	0.0018	0.0004	0	0	0	0	0	0.0008	0.002	0.0125	0.0014	0	0	0	0.0083	0.0749	0.0748	0.0014	0.0025
0.0296	0.0166	0.0021	0.0059	0.01	0.0027	0.0037	0.0002	0.0017	0.013	0	0	0	0	0	0	0	0	0.0008	0.0036	0.0079	0	0	0	0	0.0082	0.101	0.119	0	0.0361
0.0639	0.0402	0.0137	0.0231	0.0497	0.02	0.0063	0.0015	0.0046	0.0322	0	0.0101	0.0058	0.0032	0.0072	0.0082	0.001	0.0004	0.0167	0.0117	0.0378	0.0133	0.0063	0.0044	0.006	0.0296	0.1987	0.1982	0.0112	0.0105
0.0396	0.0144	0.0071	0.0128	0.0177	0.0057	0.0087	0.0017	0.0006	0.0093	0	0.0062	0.0031	0	0.0027	0.0014	0	0	0.0029	0.0036	0.0302	0.0099	0.0021	0.0019	0.0023	0.0155	0.0888	0.0888	0.0055	0.0067
0.0599	0.0322	0.0075	0.0158	0.0352	0.0142	0.0068	0.0008	0.0025	0.0296	0	0.0092	0.0041	0	0.0038	0.0003	0	0	0.011	0.0087	0.0374	0.0152	0.0022	0.0017	0.002	0.0166	0.1962	0.2006	0.0081	0.0058

[3.37 .. 3.3	[3.39 .. 3.4	[3.44 .. 3.5	[3.50 .. 3.5	[3.54 .. 3.5	[3.57 .. 3.6	[3.62 .. 3.6	[3.66 .. 3.6	[3.69 .. 3.7	[3.71 .. 3.7	[3.77 .. 3.8	[3.81 .. 3.8	[3.87 .. 3.9	[3.93 .. 3.9	[3.95 .. 3.9	[3.99 .. 4.0	[4.05 .. 4.1	[4.10 .. 4.1	[4.16 .. 4.2	[4.21 .. 4.2	[4.24 .. 4.2	[4.27 .. 4.3	[4.30 .. 4.3	[4.35 .. 4.4	[4.41 .. 4.4	[4.47 .. 4.5	[4.55 .. 4.6	[4.61 .. 4.6	[4.63 .. 4.6	[4.89 .. 4.9
0.0058	0.0868	0.0697	0.0182	0.0275	0.0384	0.0293	0.0203	0.0256	0.082	0.066	0.0838	0.0694	0.0222	0.0237	0.0154	0.0073	0.0164	0.0062	0.0022	0.003	0.005	0.0053	0.0027	0.0034	0.0016	0.0015	0	0.0149	0
0.0104	0.2153	0.1919	0.0422	0.0729	0.1698	0.1367	0.0844	0.0915	0.2808	0.2292	0.3204	0.2493	0.0629	0.1119	0.0642	0.0201	0.0316	0.0145	0.0059	0.006	0.0093	0.0099	0.0073	0.0073	0.0044	0.0037	0	0.0545	0.0009
0.0169	0.2096	0.1752	0.042	0.0765	0.1555	0.1255	0.0824	0.0829	0.2603	0.2112	0.3148	0.2131	0.0613	0.1013	0.0418	0.0159	0.0317	0.0119	0.0053	0.0051	0.0064	0.0078	0.0025	0.0029	0	0	0	0.0368	0
0.0048	0.0984	0.0826	0.0163	0.028	0.0715	0.0637	0.0401	0.0401	0.1212	0.1086	0.151	0.1282	0.0462	0.0661	0.0417	0.0162	0.0259	0.0102	0.0043	0.0048	0.0076	0.0069	0.004	0.0033	0.0002	0	0	0.0135	0.006
0.004	0.0503	0.0445	0.0126	0.019	0.0106	0.0082	0.0072	0.0122	0.0451	0.0351	0.1192	0.0395	0.0048	0.0098	0.0076	0.006	0.063	0.0041	0.0008	0.002	0.0029	0.0024	0.002	0.0022	0.0012	0.0029	0.0012	0.0141	0
0.0151	0.1714	0.1733	0.0432	0.0575	0.079	0.0621	0.0402	0.0624	0.1997	0.1392	0.2255	0.1487	0.0294	0.0454	0.0213	0.0095	0.0333	0.0066	0.0022	0.0034	0.0051	0.0053	0.0037	0.005	0.0014	0.0035	0.0014	0.0425	0
0.0151	0.1343	0.1186	0.0319	0.0513	0.0437	0.0308	0.0245	0.0358	0.1361	0.1082	0.1168	0.1062	0.0199	0.0379	0.0239	0.0135	0.0321	0.0118	0.005	0.0079	0.0085	0.0091	0.0074	0.0069	0.0046	0.0045	0.0015	0.0237	0
0.016	0.1727	0.1445	0.0379	0.0587	0.0507	0.0347	0.0276	0.0412	0.1553	0.1376	0.1334	0.1242	0.0223	0.04	0.0311	0.0155	0.0358	0.0171	0.0072	0.0099	0.0116	0.0126	0.0108	0.011	0.0083	0.0137	0.0018	0.0327	0
0.0102	0.1362	0.1328	0.0337	0.0532	0.0484	0.0367	0.0245	0.0395	0.1394	0.1004	0.1316	0.1055	0.0203	0.0317	0.0199	0.01	0.0216	0.0078	0.0029	0.0041	0.0049	0.005	0.0042	0.004	0.001	0.0021	0	0.0279	0
0.0068	0.0814	0.0861	0.0221	0.0302	0.0262	0.0197	0.0151	0.0272	0.0897	0.0616	0.0858	0.0675	0.0122	0.0201	0.0131	0.0059	0.0123	0.0041	0.002	0.0032	0.004	0.0043	0.0028	0.0031	0.0009	0.0012	0	0.021	0
0.0083	0.1212	0.11	0.0267	0.0373	0.039	0.0287	0.02	0.034	0.1123	0.0859	0.1081	0.0894	0.0161	0.0242	0.0176	0.01	0.0248	0.0088	0.0036	0.0046	0.0057	0.0056	0.0048	0.0059	0.0028	0.0047	0.0007	0.0283	0
0.0167	0.25	0.203	0.0475	0.0648	0.0592	0.0423	0.0307	0.0583	0.1969	0.162	0.1843	0.1511	0.0246	0.034	0.0249	0.0135	0.0403	0.0113	0.0054	0.0065	0.0096	0.0092	0.0084	0.0117	0.0091	0.0185	0.0047	0.0573	0
0.0193	0.2551	0.2349	0.0549	0.0794	0.1117	0.0835	0.0542	0.0833	0.2666	0.1984	0.2652	0.2064	0.0385	0.0672	0.0304	0.0145	0.0375	0.0116	0.0034	0.0048	0.0062	0.0068	0.0047	0.0058	0.0025	0.0069	0.0007	0.0504	0
0.0114	0.1437	0.124	0.0304	0.0442	0.0261	0.0201	0.0148	0.0311	0.1163	0.0966	0.1048	0.0945	0.0114	0.0208	0.0139	0.009	0.0245	0.0083	0.0024	0.0045	0.0057	0.0048	0.003	0.0048	0.0011	0.0061	0.0007	0.0238	0
0.0059	0.0738	0.06	0.0125	0.0236	0.0552	0.049	0.0305	0.0299	0.0937	0.0793	0.1177	0.0833	0.024	0.044	0.0222	0.008	0.016	0.0051	0.002	0.003	0.0032	0.0037	0.001	0.0005	0	0	0	0.0115	0.0018
0.0106	0.1022	0.1062	0.0283	0.0485	0.0418	0.0262	0.0217	0.0303	0.1242	0.0934	0.1145	0.1028	0.0226	0.0435	0.0259	0.0127	0.0268	0.0103	0.0032	0.0069	0.0071	0.0066	0.0033	0.003	0.0001	0	0	0.0154	0.0021
0.0124	0.1462	0.1142	0.0339	0.0639	0.0696	0.0522	0.0369	0.0432	0.1419	0.1241	0.14	0.1236	0.0318	0.0509	0.0334	0.0176	0.0342	0.0128	0.0047	0.0068	0.0082	0.0088	0.0059	0.0096	0.0047	0.0011	0.0001	0.0193	0
0.0132	0.1503	0.1334	0.033	0.0668	0.1055	0.0849	0.0573	0.0591	0.1966	0.1641	0.2206	0.1766	0.054	0.0834	0.0391	0.0159	0.0341	0.0112	0.0048	0.0078	0.0089	0.0108	0.0061	0.0056	0.003	0.0022	0	0.0276	0.0008
0.0109	0.0996	0.0943	0.0239	0.0512	0.0744	0.0551	0.0398	0.0415	0.1445	0.1243	0.1511	0.1289	0.0391	0.0634	0.0343	0.0141	0.0268	0.0104	0.004	0.0069	0.0073	0.0087	0.0051	0.0049	0.0006	0	0	0.0151	0.0027
0.011	0.1939	0.1559	0.0375	0.0677	0.1682	0.1413	0.0814	0.0773	0.2285	0.2015	0.3148	0.1925	0.0547	0.1025	0.0444	0.018	0.0299	0.0207	0.0073	0.0075	0.0106	0.0104	0.0095	0.011	0.0073	0.0085	0.001	0.0446	0
0.0089	0.1765	0.1613	0.037	0.0526	0.0897	0.0729	0.0442	0.0574	0.1865	0.1469	0.2158	0.1523	0.0352	0.0596	0.03	0.0122	0.0252	0.0131	0.0049	0.0052	0.007	0.0061	0.0046	0.0044	0.0004	0	0	0.0341	0.0015
0.0101	0.148	0.1239	0.032	0.0494	0.0323	0.0294	0.0214	0.0332	0.1177	0.1016	0.107	0.0879	0.0142	0.022	0.0208	0.0112	0.0227	0.01	0.0054	0.0068	0.0093	0.0085	0.0085	0.0105	0.0066	0.0089	0.0021	0.031	0
0.0134	0.196	0.1727	0.0406	0.068	0.0873	0.0706	0.0458	0.0608	0.2026	0.1573	0.2022	0.1499	0.0316	0.0519	0.0292	0.0121	0.0249	0.0095	0.0043	0.0056	0.0063	0.0061	0.005	0.005	0.0017	0.0027	0	0.0346	0
0.0089	0.1848	0.1656	0.0371	0.0636	0.0799	0.0599	0.0382	0.054	0.1913	0.1547	0.1964	0.1489	0.0321	0.0564	0.0306	0.0112	0.0234	0.008	0.0034	0.0048	0.0055	0.0053	0.0037	0.0027	0	0.0001	0	0.0298	0
0.0092	0.1174	0.1119	0.0309	0.0456	0.0283	0.0251	0.0173	0.0299	0.1092	0.0807	0.1	0.0839	0.0129	0.0248	0.0183	0.0126	0.0264	0.009	0.0043	0.0054	0.0065	0.0063	0.0047	0.0035	0.0012	0.0001	0	0.0182	0
0.0115	0.1449	0.1384	0.0352	0.0523	0.0649	0.0495	0.035	0.0521	0.1663	0.1173	0.1614	0.1272	0.0263	0.0523	0.0251	0.0115	0.026	0.0074	0.0035	0.0048	0.0058	0.0068	0.0052	0.0059	0.0045	0.0071	0.0028	0.0401	0.0727
0.0084	0.1462	0.1266	0.029	0.0498	0.0752	0.0622	0.0416	0.0517	0.1659	0.1244	0.1799	0.139	0.0382	0.0814	0.0282	0.0114	0.0213	0.0091	0.0044	0.0048	0.006	0.0078	0.0042	0.0043	0.0016	0.0021	0	0.0252	0
0.0217	0.2231	0.1814	0.0463	0.0891	0.1973	0.1572	0.0982	0.0946	0.298	0.2367	0.356	0.2451	0.0681	0.124	0.0482	0.0212	0.0407	0.025	0.0072	0.0099	0.0111	0.0128	0.0088	0.008	0.004	0.0047	0.0018	0.0475	0
0.0121	0.1704	0.1412	0.0316	0.0565	0.1227	0.1133	0.073	0.0683	0.2114	0.1887	0.2776	0.1948	0.06	0.1085	0.0585	0.0229	0.0446	0.0173	0.0078	0.0102	0.0123	0.0153	0.0107	0.0092	0.0058	0.0006	0	0.032	0.0058
0.0103	0.1325	0.1194	0.0273	0.0589	0.1234	0.1009	0.0653	0.062	0.2064	0.1716	0.254	0.1797	0.0525	0.1003	0.047	0.0189	0.0365	0.0131	0.005	0.0079	0.0088	0.0103	0.006	0.0052	0.0021	0	0	0.0282	0
0.0082	0.1644	0.1556	0.0368	0.059	0.1083	0.0886	0.0557	0.0645	0.2062	0.1618	0.2342	0.188	0.045	0.078	0.0437	0.0195	0.0345	0.0111	0.0049	0.0053	0.0071	0.0092	0.0053	0.0062	0.0005	0	0	0.0313	0.0048
0.0102	0.1898	0.1708	0.0356	0.0653	0.1703	0.1493	0.0874	0.0872	0.2744	0.2253	0.3625	0.2316	0.0646	0.1244	0.0513	0.0166	0.0282	0.015	0.0052	0.0068	0.0078	0.0083	0.0061	0.0051	0.0014	0	0	0.0421	0.0007
0.0094	0.165	0.1466	0.0294	0.0502	0.1631	0.1542	0.0901	0.0835	0.2449	0.2204	0.3602	0.2471	0.0657	0.1351	0.0594	0.0185	0.039	0.0183	0.0068	0.0068	0.0101	0.0109	0.0072	0.0065	0.0016	0	0	0.0395	0.0057
0.0062	0.1245	0.1145	0.0302	0.0445	0.0473	0.0413	0.0306	0.0391	0.144	0.3544	0.3609	0.1435	0.0555	0.0391	0.0297	0.0173	0.0237	0.0082	0.0051	0.0043	0.0069	0.0064	0.0023	0.0029	0.0014	0	0	0.0195	0
0.0137	0.2173	0.2175	0.0506	0.0759	0.1449	0.1229	0.0726	0.0897	0.2825	0.2121	0.3369	0.221	0.0534	0.0937	0.0402	0.0134	0.0229	0.0078	0.0038	0.0045	0.0062	0.0055	0.0041	0.0054	0.0007	0.001	0	0.0519	0
0.0206	0.2349	0.2208	0.0548	0.0875	0.1192	0.0905	0.0604	0.0798	0.2738	0.213	0.2706	0.2033	0.0454	0.0808	0.0439	0.0189	0.0407	0.0149	0.0067	0.009	0.0106	0.0105	0.0084	0.0077	0.0045	0.0041	0.0008	0.0509	0
0.006	0.108	0.1114	0.0266	0.0406	0.0489	0.0355	0.0241	0.0363	0.1336	0.1015	0.135	0.1104	0.0258	0.0446	0.0273	0.0106	0.0211	0.0093	0.0039	0.0053	0.0062	0.0059	0.004	0.0034	0	0	0	0.019	0.0014
0.0101	0.1769	0.1862	0.044	0.0593	0.073	0.0581	0.0373	0.06	0.2026	0.1457	0.2106	0.1586	0.0339	0.0506	0.0301	0.0126	0.0276	0.0077	0.004	0.0039	0.0056	0.0055	0.0029	0.003	0	0	0	0.0363	0.0021
0.0053	0.076	0.0822	0.0198	0.0281	0.0444	0.0355	0.0227	0.0314	0.1017	0.0696	0.1687	0.0913	0.0193	0.0361	0.0169	0.0062	0.0502	0.0075	0.003	0.0046	0.0065	0.0057	0.0042	0.0066	0.0054	0.0008	0	0.0194	0.0011
0.0217	0.3483	0.3394	0.0873	0.1198	0.1307	0.1237	0.0767	0.1087	0.3529	0.2918	0.4143	0.2967	0.0616	0.0931	0.0605	0.0251	0.0555	0.0152	0.0121	0.0123	0.0202	0.0148	0.0161	0.0197	0.0114	0.0162	0	0.0783	0
0.0149	0.1633	0.1759	0.0482	0.0729	0.076	0.0593	0.0404	0.0617	0.209	0.1696	0.2066	0.1553	0.0383	0.0467	0.0291	0.0149	0.0352	0.0062	0.0043	0.0052	0.0079	0.0072	0.0057	0.0073	0.0023	0.0042	0	0.0342	0
0.008	0.0907	0.1032	0.0291	0.0417	0.046	0.0381	0.0293	0.0398	0.1263	0.0916	0.1317	0.1142	0.0355	0.052	0.0278	0.028	0.0241	0.0074	0.0041	0.0041	0.0048	0.0047	0.002	0.0091	0	0	0	0.0185	0
0.0143	0.164	0.1552	0.0384	0.0562	0.0574	0.0426	0.0297	0.0498	0.1712	0.1273	0.1566	0.1206	0.0213	0.0311	0.0175	0.007	0.0185	0.0034	0.0021	0.003	0.0045	0.0041	0.0031	0.0042	0.0033	0.0073	0.0023	0.0358	0
0.0206	0.2891	0.2721	0.0639	0.1004	0.1463	0.1143	0.0722	0.0988	0.3256	0.2454	0.3398	0.2409	0.0531	0.0833	0.0436	0.0151	0.0367	0.0109	0.0051	0.0061	0.0075	0.0076	0.0058	0.0067	0.0024	0.004	0	0.053	0
0.0217	0.3129	0.3076	0.0724	0.1053	0.1426	0.1151	0.0708	0.1074	0.3518	0.2542	0.3757	0.2611	0.0551	0.0863	0.0427	0.0159	0.0432	0.0091	0.0045	0.0059	0.007	0.0063	0.0039	0.0055	0	0.0005	0	0.0608	0
0.0098	0.1356	0.128	0.0316	0.0462	0.0465	0.0352	0.024	0.04	0.1395	0.1027	0.1318	0.1	0.0192	0.0287	0.0183	0.0073	0.0166	0.0048	0.0026	0.0031	0.0036	0.003	0.0025	0.0024	0.0001	0.0012	0	0.023	0
0.0192	0.2467	0.2446	0.062	0.0804	0.0863	0.0631	0.0417	0.0732	0.247	0.1751	0.2377	0.1819	0.0341	0.0473	0.0284	0.013	0.0353	0.0089	0.0043	0.0058	0.0082	0.0085	0.0083	0.0114	0.0106	0.0189	0.0066	0.0698	0.0253
0.0156	0.1569	0.1315	0.0334	0.0719	0.1147	0.0841	0.0575	0.0594	0.2085	0.1696	0.2124	0.1544	0.0423	0.0796	0.0366	0.0133	0.0282	0.0105	0.0041	0.0082	0.008	0.0082	0.0052	0.0049	0.0023	0.0015	0.0002	0.0258	0
0.0085	0.1311	0.1219	0.0285	0.0513	0.0768	0.064	0.0454	0.0495	0.168	0.1254	0.1902	0.2872	0.0702	0.0726	0.0348	0.0145	0.0291	0.0116	0.0046	0.0063	0.0075	0.0085	0.0043	0.0039	0.002	0	0	0.0219	0.0021
0.009	0.136	0.1287	0.0285	0.0551	0.1156	0.0956	0.0592	0.0611	0.2019	0.1624	0.2445	0.1615	0.0452	0.0855	0.0365	0.0136	0.0255	0.0101	0.0038	0.0059	0.0061	0.0064	0.0041	0.0036	0.0014	0	0	0.0281	0.0005
0.0144	0.1782	0.153	0.0367	0.0702	0.1365	0.1183	0.0732	0.0728	0.2335	0.1946	0.2854	0.1911	0.0587	0.0978	0.0413	0.0172	0.0332	0.0108	0.0051	0.0069	0.008	0.0082	0.0055	0.0059	0.0034	0.0047	0.0011	0.0342	0.0012
0.0128	0.1381	0.1218	0.0323	0.0615	0.0507	0.0372	0.0275	0.0361	0.1448	0.1156	0.1316	0.1094	0.0238	0.0451	0.026	0.0132	0.0288	0.0112	0.0046	0.0078	0.0083	0.007	0.0055	0.0052	0.0019	0.0016	0	0.0194	0
0.0151	0.1447	0.1377	0.0411	0.0727	0.0519	0.0365	0.0289	0.0387	0.1515	0.119	0.1375	0.1288	0.0269	0.0524	0.0313	0.0174	0.0381	0.0116	0.0051	0.0089	0.0095	0.009	0.0063	0.0054	0.0034	0.0013	0.0001	0.0215	0
0.0201	0.2708	0.2414	0.0668	0.1157	0.0664	0.0589	0.0404	0.0571	0.2186	0.1865	0.2098	0.1678	0.0289	0.0551	0.0472	0.024	0.0532	0.0239	0.0121	0.0152	0.016	0.0175	0.0159	0.0129	0.0088	0.0078	0.0003	0.05	0.0001
0.0297	0.3624	0.3435	0.0826	0.1214	0.1244	0.0981	0.0649	0.1082	0.3643	0.2607	0.3407	0.2544	0.0438	0.0656	0.0392	0.0153	0.0405	0.0098	0.0048	0.0063	0.0074	0.0066	0.005	0.006	0.0009	0.0049	0	0.0653	0
0.0175	0.2278	0.2143	0.0531	0.0781	0.0708	0.0552	0.0364	0.0635	0.2234	0.1604	0.2066	0.1581	0.0269	0.0393	0.0249	0.0114	0.0284	0.0087	0.0037	0.0045	0.0059	0.0054	0.0033	0.004	0	0.0016	0	0.0376	0
0.0085	0.1473	0.1305	0.0346	0.0616	0.0417	0.0374	0.0232	0.0334	0.1324	0.1142	0.1248	0.1028	0.0201	0.0345	0.0242	0.0129	0.0245	0.0102	0.0051	0.0062	0.007	0.0075	0.0058	0.0045	0.0005	0.0007	0	0.019	0.0004
0.005	0.1151	0.1195	0.0294	0.0424	0.0426	0.0308	0.0199	0.0341	0.1268	0.0961	0.1396	0.1073	0.023	0.0341	0.0237	0.011	0.0278	0.0092	0.004	0.0044	0.0053	0.005	0.0039	0.0036	0.001	0	0	0.0212	0.0038
0.0181	0.184	0.1785	0.0481	0.0745	0.0385	0.0351	0.0239	0.0452	0.1697	0.1133	0.1476	0.1173	0.0143	0.0264	0.0153	0.0124	0.0322	0.0092	0.0037	0.0054	0.0069	0.0067	0.004	0.0039	0.0015	0.0024	0.0012	0.0354	0
0.0163	0.2134	0.2002	0.0527	0.0838	0.0583	0.0498	0.036	0.0583	0.2063	0.1409	0.1852	0.1614	0.0329	0.0778	0.0318	0.0167	0.0403	0.0144	0.007	0.0092	0.0113	0.0128	0.009	0.0095	0.0035	0.0043	0.0005	0.0387	0.0005
0.0065	0.1289	0.1366	0.0332	0.0498	0.0402	0.0291	0.0194	0.0364	0.1391	0.1026	0.1337	0.1084	0.0203	0.0346	0.0223	0.0088	0.0218	0.0066	0.0027	0.0041	0.0045	0.0038	0.002	0.002	0	0	0	0.0224	0.0002
0.0099	0.1979	0.1971	0.0458	0.0609	0.0937	0.0799	0.0476	0.0683	0.2262	0.1743	0.265	0.1844	0.0424	0.0678	0.0355	0.0123	0.0275	0.0105	0.0047	0.0052	0.0081	0.0065	0.0045	0.0051	0	0	0	0.0442	0.0026
0.0292	0.3692	0.3548	0.0871	0.1334	0.1342	0.1068	0.0704	0.1119	0.3864	0.2711	0.3578	0.2676	0.0484	0.0708	0.0426	0.0158	0.0404	0.0076	0.0046	0.0063	0.0068	0.0065	0.004	0.004	0	0.0012	0	0.0656	0
0.0128	0.172	0.1631	0.0386	0.0686	0.1224	0.0954	0.0607	0.0688	0.229	0.1773	0.2527	0.1754	0.0436	0.0839	0.037	0.0117	0.0211	0.0068	0.0029	0.0055	0.0052	0.0048	0.0025	0.0015	0	0	0	0.0305	0
0.022	0.3283	0.3149	0.0724	0.1149	0.1914	0.1472	0.0938	0.1185	0.3978	0.3061	0.4323	0.3118	0.0716	0.1255	0.0605	0.0214	0.05	0.0182	0.0064	0.0094	0.0101	0.0109	0.0067	0.0057	0.0021	0.003	0	0.0639	0

[5.17 .. 5.2	[5.20 .. 5.2	[5.22 .. 5.2	[5.38 .. 5.4	[5.88 .. 5.9	[6.09 .. 6.1	[6.49 .. 6.5	[6.78 .. 6.8	[6.81 .. 6.8	[6.84 .. 6.8	[6.86 .. 6.8	[6.88 .. 6.9	[7.04 .. 7.0	[7.07 .. 7.1	[7.12 .. 7.1	[7.15 .. 7.1	[7.17 .. 7.2	[7.23 .. 7.2	[7.28 .. 7.3	[7.30 .. 7.3	[7.35 .. 7.4	[7.41 .. 7.4	[7.47 .. 7.5	[7.52 .. 7.5	[7.58 .. 7.6	[7.62 .. 7.6	[7.65 .. 7.7	[7.70 .. 7.7	[7.72 .. 7.7	[7.78 .. 7.8
0.0004	0.0002	0.0186	0.0158	0.0003	0.0005	0.0004	0.0001	0.0002	0.0001	0.0002	0.0016	0.0002	0.0007	0.0007	0.0011	0.0031	0.001	0.0004	0.0021	0.002	0.0018	0.0005	0.001	0.0012	0.0002	0.0005	0.0002	0.0006	0.0004
0.0016	0	0.0654	0.1033	0.0015	0.0003	0	0	0.0001	0	0	0.002	0.0001	0.0009	0.0004	0.0016	0.0039	0.0017	0.001	0.0042	0.0025	0.0026	0.0005	0.001	0.0021	0.0001	0.0006	0.0002	0.0008	0.0005
0.0017	0.0001	0.0553	0.0895	0.0007	0	0.0002	0.0002	0.0003	0.0004	0	0.0047	0.0005	0.0042	0.0008	0.0007	0.0079	0.0022	0.0012	0.0075	0.0043	0.0054	0.0007	0.0023	0.0023	0.0004	0.0011	0.0005	0.0014	0.0011
0.0001	0	0.0291	0.0503	0.0004	0.0002	0	0	0.0001	0	0	0.0007	0	0	0	0.0012	0.0041	0.0009	0.0005	0.003	0.0022	0.0023	0.0002	0.0003	0.0006	0.0001	0.0001	0	0.0001	0
0	0	0.0077	0	0	0	0.0007	0.0006	0.0012	0.001	0.0013	0.0042	0.0297	0.0281	0.0009	0.0009	0.0038	0.0032	0.0017	0.006	0.0093	0.055	0.0005	0.0012	0.0014	0.0002	0.0018	0.0005	0.0018	0.002
0.0025	0.0004	0.0511	0.0403	0	0	0	0.0004	0.0002	0.0004	0.0004	0.0027	0.0071	0.0089	0.0009	0.0003	0.0044	0.0025	0.0011	0.0048	0.0043	0.0155	0.0013	0.0015	0.0017	0.0003	0.001	0.0003	0.0011	0.001
0.002	0.0008	0.0283	0.0071	0.0009	0	0.0009	0.0007	0.0007	0.0006	0.0006	0.0064	0.0016	0.005	0.0015	0.0014	0.0094	0.0031	0.0014	0.0095	0.0055	0.0071	0.0015	0.0028	0.0021	0.0005	0.0006	0.0003	0.0016	0.0042
0.0017	0.0008	0.0327	0.0072	0.0012	0.0007	0.0005	0.001	0.001	0.0007	0.0006	0.0067	0.001	0.003	0.0018	0.0042	0.0098	0.0036	0.0015	0.0088	0.0053	0.0059	0.0004	0.0017	0.0019	0.0003	0.0007	0.0001	0.0005	0.001
0.0007	0	0.036	0.0166	0.0007	0	0	0.0003	0.0003	0.0002	0.0003	0.0033	0.0001	0.0019	0.0006	0.0004	0.0045	0.0014	0.0006	0.0045	0.003	0.0038	0.0009	0.0016	0.0021	0.0006	0.0008	0.0001	0.001	0.0006
0.0003	0	0.023	0.0092	0	0	0	0.0006	0.0006	0.0004	0.0004	0.0027	0.0006	0.0017	0.0007	0.0017	0.0046	0.0025	0.0012	0.0048	0.0034	0.0037	0.0011	0.0021	0.0024	0.0005	0.0014	0.0006	0.0015	0.0014
0.0004	0	0.0285	0.0149	0.0006	0.0001	0	0.0007	0.0005	0.0002	0.0005	0.0024	0.0002	0.0009	0.0005	0.0012	0.0038	0.0017	0.001	0.0037	0.0021	0.0024	0.001	0.0014	0.0021	0.0003	0.0007	0.0006	0.0009	0.0002
0.0018	0	0.0473	0.016	0.0007	0.0001	0.0001	0.0013	0.001	0.0009	0.0008	0.0051	0.0014	0.0039	0.002	0.0036	0.0077	0.0043	0.0018	0.0077	0.0053	0.0053	0.0024	0.0037	0.0052	0.0011	0.0024	0.0011	0.0022	0.0022
0.0028	0.0006	0.0662	0.0544	0.0004	0	0.0005	0.0008	0.0006	0.0009	0.0006	0.0039	0.0009	0.0045	0.002	0.0006	0.0061	0.0024	0.0009	0.0059	0.0037	0.0071	0.0014	0.0024	0.0028	0.0003	0.0018	0.0012	0.0015	0.001
0.0006	0	0.0257	0	0	0	0.0004	0.0012	0.0008	0.0008	0.0006	0.0042	0.001	0.0039	0.001	0.0007	0.0062	0.0025	0.0012	0.0065	0.0038	0.0047	0.0017	0.0025	0.0035	0.0007	0.0014	0.0007	0.0015	0.0015
0.0003	0	0.0216	0.0384	0	0	0	0	0	0	0	0.0013	0	0.0003	0.0003	0.0009	0.0026	0.0012	0.0003	0.0027	0.0016	0.0017	0.0001	0	0.0005	0	0	0	0.0002	0
0.0014	0	0.03	0.0107	0	0	0	0	0	0	0	0.0059	0	0.0027	0.0003	0.0003	0.0088	0.0018	0.0009	0.0095	0.0053	0.009	0.0003	0.0017	0.0016	0.0002	0.0002	0.0001	0.001	0
0.0015	0.0002	0.0251	0.0243	0.0003	0.0001	0	0.0002	0.0005	0.0005	0.0003	0.0049	0.0004	0.0016	0.0009	0.003	0.008	0.0023	0.0012	0.0063	0.0039	0.0048	0.0009	0.0019	0.0021	0	0.001	0.0002	0.0013	0.0008
0.0028	0.0002	0.0429	0.0586	0.0009	0	0.0006	0.0001	0.0003	0.0002	0.0005	0.0064	0.0002	0.0031	0.0007	0.0002	0.0087	0.0019	0.0009	0.0091	0.0058	0.0112	0.0088	0.0054	0.0019	0.0004	0.0003	0.0001	0.001	0.0014
0.0019	0	0.0312	0.0351	0.0007	0	0	0	0	0	0	0.0051	0	0.0018	0	0.0001	0.007	0.0009	0.0005	0.0081	0.0039	0.0078	0.0024	0.0017	0.0012	0	0	0	0.0004	0
0.0023	0.0001	0.0532	0.1113	0.0022	0.0016	0	0.0007	0.0002	0.0003	0.0005	0.0028	0.0011	0.0022	0.0007	0.002	0.0045	0.0018	0.0006	0.0044	0.003	0.0038	0.0005	0.0007	0.0018	0.0001	0.0005	0	0.0002	0
0.0013	0	0.0502	0.0581	0.0017	0.0008	0	0.0001	0.0004	0.0003	0.0002	0.0022	0	0.0007	0.0006	0.0015	0.0039	0.0011	0.0002	0.0035	0.0023	0.0027	0.0004	0.0003	0.0019	0.0003	0.0003	0.0001	0.0004	0.0001
0.0009	0	0.0285	0.0054	0.0023	0.002	0.0004	0.0015	0.0012	0.0009	0.0007	0.0039	0.001	0.0024	0.0013	0.0023	0.0057	0.0031	0.0009	0.0062	0.0036	0.0037	0.0017	0.002	0.0034	0.0006	0.001	0.0008	0.002	0.001
0.0029	0.0008	0.0508	0.0414	0.0006	0	0	0.0007	0.0001	0.0004	0.0005	0.0037	0.001	0.0032	0.0007	0.0006	0.0054	0.0015	0.001	0.0061	0.0038	0.0055	0.0011	0.0023	0.003	0.0006	0.0007	0.0003	0.0012	0.0008
0.0009	0	0.0485	0.0402	0.0014	0	0	0.0003	0	0	0	0.0027	0	0.0025	0.0003	0.0001	0.0044	0.0011	0.0005	0.006	0.0037	0.0055	0.0007	0.0022	0.0025	0.0003	0.0004	0.0002	0.0007	0.0004
0.001	0.0001	0.0273	0.0036	0	0	0	0.0008	0.0006	0.0004	0.0005	0.0034	0.0007	0.0034	0.0008	0.0009	0.0057	0.0025	0.0014	0.0064	0.0043	0.0048	0.0019	0.003	0.0027	0.0003	0.0006	0.0004	0.0012	0.001
0.0011	0	0.0369	0.022	0	0	0	0	0	0.0001	0.0001	0.003	0.0016	0.0039	0.0017	0.0008	0.0061	0.0023	0.0014	0.0061	0.0045	0.0078	0.0019	0.0045	0.0026	0.0017	0.0018	0.0009	0.0036	0.0032
0.0018	0	0.0369	0.0423	0.0008	0	0	0.0007	0.0004	0.0003	0.0002	0.0033	0.0007	0.0042	0.0023	0.002	0.0083	0.0024	0.001	0.0063	0.0042	0.0105	0.0027	0.0063	0.0036	0.0041	0.0029	0.0022	0.0057	0.0053
0.0031	0.0007	0.06	0.1071	0.0008	0	0.0002	0.0006	0.0006	0.0005	0.0002	0.0072	0.0008	0.0054	0.0012	0.0011	0.0111	0.0033	0.0017	0.0133	0.0075	0.0093	0.0024	0.0034	0.0023	0.0005	0.0007	0.0006	0.0015	0.0013
0.0029	0	0.0495	0.0891	0.0014	0	0	0	0.0001	0	0.0001	0.0042	0	0.0001	0.0006	0.003	0.0069	0.0018	0.0007	0.0077	0.0044	0.0063	0.0011	0.0011	0.0014	0.0005	0.0002	0.0002	0.0004	0
0.0007	0	0.0436	0.0794	0.0002	0	0	0	0	0	0	0.0057	0	0.0029	0.0005	0.0003	0.0081	0.0016	0.0007	0.0099	0.0058	0.0084	0.0007	0.002	0.0016	0.0004	0	0.0002	0.0007	0
0.0019	0	0.0532	0.0669	0.0023	0	0.0007	0	0	0	0	0.0019	0	0.0016	0.0001	0	0.0032	0.0011	0.0003	0.0039	0.0016	0.0028	0	0.0006	0.0014	0	0.0001	0	0.0002	0
0.002	0	0.0639	0.1265	0.0014	0	0	0	0.0001	0	0	0.004	0	0.0027	0.0004	0.0003	0.0062	0.0014	0.0006	0.0063	0.0032	0.0048	0.0003	0.0014	0.0013	0.0002	0	0.0002	0.0008	0.0004
0.0016	0	0.058	0.1326	0.0017	0	0	0	0	0	0	0.0019	0.0037	0.0063	0.001	0.0019	0.0061	0.0016	0.0008	0.004	0.0018	0.0144	0.0011	0.0015	0.0012	0.0001	0.001	0.0006	0.0006	0.0006
0.0013	0	0.0303	0.0173	0.0002	0	0	0	0.0003	0.0003	0.0029	0.076	0.0021	0.0084	0.0009	0.0005	0.0081	0.0751	0.0043	0.0297	0.1852	0.0493	0.0142	0.0808	0.0157	0.005	0.0042	0.0039	0.009	0.0038
0.0025	0	0.0727	0.0976	0.0005	0	0	0.0004	0.0003	0.0002	0.0003	0.0029	0.0003	0.0026	0.0005	0.0007	0.0054	0.0019	0.0011	0.0055	0.0032	0.0045	0.0009	0.0015	0.0018	0.0006	0.0005	0.0001	0.0008	0.0008
0.0029	0.0006	0.0638	0.0556	0.0015	0	0	0.0006	0.0004	0.0004	0.0004	0.0072	0.0003	0.0043	0.001	0.0008	0.0109	0.0026	0.0015	0.011	0.0057	0.0079	0.0003	0.0015	0.0017	0.0005	0.0004	0.0001	0.0011	0.0007
0.0009	0	0.0325	0.0223	0	0	0	0	0	0	0	0.0037	0.0001	0.0019	0.0004	0.0004	0.0062	0.0018	0.0005	0.0064	0.004	0.0055	0.0013	0.0013	0.0015	0.0002	0.0006	0.0002	0.0007	0.0001
0.0025	0	0.0578	0.0421	0.0008	0	0	0.0007	0.0004	0.0004	0.0003	0.0026	0.0004	0.0022	0.0006	0.0005	0.0038	0.0015	0.0006	0.0041	0.0021	0.003	0.0006	0.0006	0.0016	0	0.0004	0	0.0003	0
0.001	0	0.0264	0.0309	0.0008	0.0001	0	0	0	0	0	0.0009	0.0177	0.0252	0.0005	0.0001	0.0031	0.001	0.0007	0.0026	0.0018	0.0421	0.0023	0.0014	0.0014	0.0007	0.0011	0.0003	0.0011	0.0013
0.0054	0	0.0905	0.0788	0	0.0029	0	0.0025	0.0014	0.0013	0.001	0.0049	0.0025	0.0049	0.0028	0.0047	0.0087	0.0057	0.0023	0.0074	0.0053	0.0055	0.0036	0.0032	0.0051	0.0012	0.0025	0.0012	0.0022	0.0022
0.003	0	0.0508	0.0288	0.0015	0	0	0.0013	0.0007	0.0006	0.0004	0.0044	0.0007	0.0034	0.0008	0.0006	0.0066	0.0026	0.0011	0.0066	0.0041	0.0059	0.0018	0.0021	0.0032	0.0005	0.0012	0.0007	0.001	0.0015
0.0007	0	0.0275	0.0195	0	0	0	0.0001	0.0001	0.0001	0.0001	0.0018	0.0007	0.0033	0.003	0.0028	0.0054	0.0029	0.0012	0.0049	0.0038	0.0071	0.0055	0.0076	0.0074	0.0018	0.0048	0.0027	0.0087	0.0065
0.0017	0.0001	0.0391	0.0173	0	0	0	0.0007	0.0003	0.0002	0.0004	0.0034	0.0002	0.0027	0.001	0.001	0.0063	0.0022	0.0013	0.006	0.0033	0.0043	0.0006	0.0019	0.002	0.0004	0.0013	0.0003	0.0014	0.0014
0.0047	0	0.0808	0.077	0.0023	0	0.0004	0.0014	0.0003	0.0006	0.0006	0.0047	0.0009	0.0043	0.0011	0.001	0.0074	0.0027	0.0011	0.0074	0.0045	0.0066	0.0012	0.002	0.0026	0.0003	0.0009	0.0002	0.0015	0.0007
0.0041	0	0.0915	0.0795	0.0017	0	0	0.0008	0.0003	0.0002	0.0002	0.0046	0.0003	0.0038	0.0011	0.0007	0.0075	0.0027	0.001	0.0079	0.0048	0.0063	0.0013	0.0023	0.0031	0.0005	0.0015	0.0004	0.0023	0.0009
0.0019	0	0.0363	0.0193	0.0015	0	0	0.0005	0.0001	0.0003	0.0002	0.0029	0.0003	0.0021	0.0004	0.0004	0.0046	0.001	0.0007	0.0043	0.0025	0.0031	0.0005	0.0011	0.0015	0.0001	0	0.0001	0.0008	0.0002
0.0064	0.0025	0.0685	0.0356	0.0003	0	0	0.0002	0.0001	0	0.0001	0.003	0	0.0024	0.0007	0.0003	0.0058	0.0013	0.0009	0.0053	0.0035	0.0041	0.0004	0.0006	0.0015	0	0.0004	0.0002	0.0005	0
0.0019	0.0001	0.0398	0.052	0.001	0	0	0.0006	0.0004	0.0002	0.0002	0.0085	0.0007	0.0051	0.0011	0.0012	0.012	0.0029	0.0016	0.0137	0.0068	0.0096	0	0.0022	0.0013	0.0003	0.0006	0.0003	0.0012	0.0003
0.0021	0	0.0383	0.0411	0.0003	0	0	0	0	0	0	0.0052	0.0016	0.0306	0.0029	0.0011	0.0075	0.0023	0.0017	0.0077	0.0076	0.0131	0.0518	0.0108	0.0099	0.0191	0.0268	0.0072	0.0311	0.0064
0.0012	0	0.0464	0.078	0.0011	0	0	0.0001	0.0002	0	0	0.0044	0	0.0021	0.0006	0.0003	0.0065	0.0017	0.0005	0.0073	0.0038	0.0065	0.0007	0.0012	0.0011	0.0001	0.0002	0.0003	0.0012	0.0001
0.0023	0.0003	0.0517	0.0872	0.001	0	0.0001	0.0002	0.0001	0	0.0001	0.0051	0.0004	0.0036	0.0005	0.0006	0.0081	0.0019	0.001	0.0083	0.0049	0.007	0.0013	0.0019	0.0013	0	0.0003	0.0005	0.0014	0.0006
0.0019	0	0.0289	0.0092	0.0017	0	0	0.0003	0.0003	0.0003	0.0005	0.0073	0.0003	0.0045	0.0006	0.0009	0.011	0.0026	0.0015	0.0119	0.0066	0.0093	0.0005	0.0025	0.0012	0.0002	0.0008	0.0004	0.0011	0.0007
0.0007	0	0.0293	0.0084	0	0	0	0.0005	0.0005	0.0003	0.0002	0.008	0.0002	0.0049	0.0007	0.0008	0.0117	0.0026	0.0013	0.0124	0.0069	0.0095	0.001	0.0023	0.0019	0	0.0004	0.0003	0.0016	0.0002
0.0033	0.0003	0.057	0.0047	0.0044	0	0.0001	0.003	0.0018	0.0013	0.0015	0.0111	0.0017	0.0041	0.0024	0.006	0.0163	0.005	0.0025	0.0162	0.0091	0.0122	0.0023	0.0038	0.0039	0.0007	0.0017	0.0007	0.0023	0.0013
0.0049	0.0007	0.0924	0.0498	0.0031	0	0	0.0019	0.0006	0.0007	0.0007	0.0067	0.0008	0.005	0.0015	0.0011	0.0096	0.0027	0.0016	0.0098	0.0052	0.0075	0.0012	0.003	0.0035	0.0008	0.0015	0.0003	0.0019	0.0013
0.0026	0	0.0544	0.0213	0.0021	0	0.0001	0.0013	0.0004	0.0003	0.0005	0.0049	0.0005	0.0037	0.0009	0.0009	0.0077	0.0026	0.0013	0.0076	0.0045	0.0059	0.0015	0.0021	0.0028	0.0008	0.0012	0.0006	0.0016	0.0011
0.0016	0	0.0347	0.0132	0.0025	0	0	0.0006	0.0003	0	0.0001	0.0046	0	0.0021	0.0003	0.0002	0.0064	0.0016	0.0006	0.0072	0.0036	0.0064	0.0006	0.0017	0.0021	0.0002	0.0006	0.0002	0.0011	0
0.0015	0	0.035	0.0193	0.0017	0	0	0.0003	0	0	0	0.0015	0.0035	0.007	0.0004	0.0002	0.003	0.0013	0.0005	0.0035	0.0016	0.0124	0.0022	0.0011	0.0015	0.0004	0.0003	0.0001	0.0005	0
0.0016	0.0004	0.0419	0.0027	0.0024	0	0.0012	0.0015	0.0006	0.0007	0.0007	0.0061	0.0011	0.0046	0.0012	0.001	0.0084	0.0031	0.0016	0.0091	0.0053	0.007	0.0015	0.0027	0.0029	0.0003	0.0011	0.0006	0.0022	0.0014
0.0038	0.0006	0.053	0.0131	0.0042	0	0.0014	0.0018	0.0008	0.0009	0.0009	0.0075	0.001	0.0073	0.0037	0.0025	0.0137	0.0034	0.0016	0.0112	0.0076	0.0162	0.0022	0.0078	0.0045	0.0052	0.004	0.0029	0.0058	0.0059
0.0006	0	0.0366	0.0141	0.0006	0	0	0.0003	0.0001	0	0	0.0032	0.0001	0.0022	0.0003	0.0002	0.0052	0.0012	0.0006	0.0057	0.0033	0.0058	0.0008	0.0015	0.0018	0.0003	0.0011	0.0003	0.0011	0
0.0025	0	0.0605	0.0649	0.0009	0	0	0.0008	0.0003	0	0	0.0018	0	0.0008	0.0009	0.0017	0.0043	0.0014	0.001	0.0043	0.0028	0.0029	0.0009	0.0008	0.002	0.0005	0.0005	0.0001	0.0001	0.0004
0.006	0.0009	0.1008	0.0543	0.0038	0	0	0.0012	0.0003	0.0001	0.0002	0.0059	0.0003	0.0044	0.001	0.0006	0.0099	0.0029	0.0016	0.0105	0.0052	0.0078	0.0012	0.0026	0.0032	0.0006	0.001	0.0006	0.0016	0.0009
0.0017	0	0.0519	0.0674	0	0	0	0.0002	0.0002	0.0002	0.0002	0.0057	0	0.0031	0.0007	0.0005	0.0095	0.0025	0.0013	0.0102	0.0064	0.0084	0.0015	0.0026	0.0018	0.0007	0.0014	0.0005	0.0018	0.0012
0.0031	0	0.0965	0.1061	0.002	0	0	0.0001	0	0.0001	0.0002	0.0084	0.0003	0.0042	0.0008	0.0008	0.0125	0.0026	0.0018	0.0125	0.0067	0.0104	0.0009	0.0029	0.0026	0.0007	0.0012	0.0003	0.002	0.001

[7.83 .. 7.8	[7.85 .. 7.9	[7.91 .. 7.9	[7.94 .. 7.9	[7.99 .. 8.0	[8.01 .. 8.0	[8.04 .. 8.0	[8.08 .. 8.1	[8.12 .. 8.1	[8.15 .. 8.1	[8.17 .. 8.2	[8.23 .. 8.2	[8.29 .. 8.3	[8.31 .. 8.3	[8.33 .. 8.3	[8.36 .. 8.3	[8.39 .. 8.4	[8.44 .. 8.4	[8.47 .. 8.5	[8.69 .. 8.7	[8.74 .. 8.7	[8.77 .. 8.8	[8.80 .. 8.8	[8.86 .. 8.9	[8.90 .. 8.9	[8.92 .. 8.9
0.0001	0.0015	0.0002	0.0014	0.0001	0.0003	0.0015	0.0007	0.0002	0.0004	0.005	0.0043	0.0004	0.0004	0.0022	0.0005	0.0009	0.0011	0.001	0.0016	0	0.0003	0.0008	0.0007	0.0003	0.0011
0.0002	0.0036	0.0002	0.0025	0.0005	0.0004	0.0025	0.0007	0.0002	0.0001	0.0084	0.0068	0.0007	0.0007	0.0028	0.0012	0.0016	0.0011	0.0013	0.0026	0.0009	0.0007	0.0009	0.0006	0.0005	0.0023
0.0009	0.0071	0.001	0.004	0.0007	0.0004	0.0009	0.0008	0.0009	0.0008	0.0086	0.0058	0.0006	0.0005	0.0024	0.0014	0.0015	0.0013	0.0015	0.003	0.0006	0.0006	0.0017	0.001	0.0007	0.0022
0	0.0006	0.0001	0.0009	0.0001	0.0001	0.002	0.0004	0	0.0003	0.004	0.0047	0.0004	0.0004	0.0025	0.0009	0.0007	0.0023	0.0015	0.001	0.0005	0.0002	0.0009	0.0008	0	0.0017
0.0003	0.0016	0.0006	0.0033	0.0007	0.0017	0.0014	0.0014	0.0006	0.0008	0.0068	0.0035	0.0009	0.0005	0.0016	0.0015	0.0017	0.0015	0.0028	0.0033	0.0006	0.0005	0.0013	0.0011	0.0009	0.0022
0.0004	0.004	0.0006	0.0032	0.0005	0.0006	0.0008	0.0008	0.0004	0.0007	0.0065	0.0047	0.0006	0.0006	0.0018	0.0011	0.0012	0.0011	0.0024	0.0029	0.0007	0.0006	0.0014	0.001	0.0005	0.0022
0.0007	0.0031	0.0038	0.0014	0.0004	0.0004	0.0009	0.0008	0.0008	0.0008	0.0097	0.0037	0.0003	0.0003	0.0012	0.0007	0.0012	0.0028	0.0016	0.0021	0.0005	0.0003	0.0012	0.0004	0.0002	0.0018
0.0002	0.0027	0.0002	0.0031	0.0006	0.0007	0.0039	0.0005	0.001	0.001	0.0086	0.0047	0.0002	0.0005	0.0024	0.0009	0.001	0.0032	0.0005	0.0021	0.0004	0.0001	0.0004	0.0007	0.0005	0.001
0.0002	0.0045	0.0007	0.0029	0.0008	0.0007	0.0011	0.0004	0.0004	0.0004	0.0062	0.0048	0.0005	0.0005	0.0015	0.0011	0.0012	0.0005	0.0014	0.0025	0.0005	0.0005	0.0011	0.0007	0.0004	0.0023
0.0008	0.0025	0.0009	0.0029	0.0006	0.0007	0.0026	0.0009	0.0009	0.0011	0.0059	0.0049	0.0007	0.0007	0.002	0.0015	0.0019	0.001	0.0014	0.0026	0.001	0.0009	0.0017	0.0011	0.0005	0.0024
0.0004	0.0028	0.0002	0.0024	0.0004	0.0007	0.0028	0.0008	0.0005	0.0007	0.0063	0.0049	0.0007	0.0004	0.0023	0.0009	0.0011	0.002	0.0013	0.0023	0.0007	0.0007	0.0013	0.0006	0.0007	0.0021
0.0011	0.0062	0.0015	0.0069	0.0019	0.002	0.005	0.0019	0.0015	0.0019	0.0174	0.0108	0.0009	0.0012	0.0056	0.0019	0.0033	0.0035	0.0036	0.0065	0.0016	0.0015	0.0029	0.0021	0.0015	0.0052
0.0006	0.0054	0.0013	0.0056	0.0007	0.0006	0.0013	0.0008	0.001	0.0009	0.0113	0.0067	0.0003	0.0006	0.0023	0.0009	0.0025	0.0014	0.002	0.0031	0.0009	0.0008	0.0015	0.0012	0.0007	0.0031
0.0008	0.0057	0.0008	0.0039	0.0009	0.0011	0.0018	0.0012	0.0014	0.0011	0.0112	0.0062	0.0007	0.0006	0.0025	0.0011	0.0017	0.0015	0.0023	0.0039	0.0008	0.0008	0.0016	0.0013	0.0007	0.0028
0	0.0005	0	0.0009	0.0002	0.0004	0.0013	0.0001	0.0002	0.0002	0.0023	0.0024	0.0002	0.0001	0.0009	0.0003	0.0005	0.0007	0.0006	0.0011	0.0004	0.0006	0.0005	0.0002	0.0003	0.0006
0.0002	0.0043	0.0004	0.0033	0.0003	0.0004	0.0007	0.0001	0.0003	0.0004	0.0044	0.0035	0.0007	0.0004	0.0009	0.0009	0.0009	0.0005	0.0006	0.0015	0	0.0001	0.0007	0.0004	0.0004	0.0019
0.0004	0.0029	0.0003	0.0034	0.0002	0.0007	0.003	0.0009	0.0007	0.0005	0.0083	0.0054	0.0009	0.0005	0.002	0.0015	0.0011	0.0005	0.001	0.0026	0.0007	0.0004	0.0012	0.0006	0.0003	0.0023
0.0014	0.0113	0.0003	0.0022	0.0002	0.0002	0.0005	0.0005	0.0002	0.0003	0.005	0.0035	0.0003	0	0.0009	0.0011	0.0007	0.0005	0.0011	0.0013	0.0005	0.0001	0.0004	0.0004	0.0001	0.0011
0.0003	0.0061	0	0.0014	0.0001	0.0002	0	0	0	0	0.0047	0.0036	0.0003	0.0002	0.0006	0.0004	0.0004	0.0004	0.0006	0.0016	0.0001	0.0003	0	0.0002	0.0003	0.0009
0	0.003	0.0001	0.0027	0.0003	0.0003	0.0023	0.0004	0	0.0002	0.0083	0.0061	0.0002	0.0004	0.0034	0.001	0.0009	0.0021	0.0009	0.003	0.0004	0.0004	0.0007	0.0004	0.0001	0.0018
0	0.0028	0.0001	0.0023	0.0005	0.0002	0.0019	0.0004	0.0002	0	0.0076	0.0062	0.0006	0.0005	0.0028	0.001	0.0012	0.0021	0.0018	0.003	0.0006	0.0004	0.0013	0.0009	0.0004	0.0025
0.0004	0.0044	0.0008	0.004	0.0008	0.0008	0.0031	0.0016	0.0011	0.0011	0.0091	0.0083	0.0006	0.0008	0.0049	0.0008	0.0017	0.0024	0.0018	0.0044	0.0008	0.0006	0.002	0.0011	0.0008	0.0035
0.0005	0.0049	0.0008	0.0042	0.0006	0.0007	0.0009	0.0012	0.0005	0.0006	0.0114	0.0057	0.0004	0.0004	0.0021	0.0009	0.0012	0.0043	0.0016	0.0033	0.0006	0.0006	0.0013	0.0009	0.0007	0.003
0	0.0055	0.0004	0.0033	0.0006	0.0004	0.0007	0.0003	0.0003	0.0004	0.007	0.0054	0.0005	0.0007	0.0018	0.0007	0.0012	0.0007	0.0021	0.0028	0.0006	0.0005	0.0011	0.0011	0.0005	0.0025
0.0003	0.0043	0.0012	0.0041	0.0007	0.0005	0.0012	0.0009	0.0007	0.0006	0.0098	0.0049	0.0005	0.0007	0.0017	0.0009	0.0013	0.0032	0.0021	0.0032	0.0003	0.0005	0.0014	0.0012	0.0006	0.0028
0.0012	0.0051	0.0014	0.0045	0.0008	0.0006	0.001	0.0012	0.0006	0.0011	0.0106	0.0062	0.0008	0.0008	0.0019	0.0011	0.0027	0.0018	0.0031	0.0032	0.0008	0.0011	0.0026	0.0015	0.0008	0.0033
0.0027	0.0071	0.002	0.0059	0.0008	0.0008	0.0009	0.0009	0.0007	0.0006	0.0116	0.0072	0.0008	0.0007	0.0021	0.0012	0.0043	0.0018	0.003	0.0035	0.0008	0.0012	0.0045	0.0009	0.0005	0.0025
0.0007	0.0071	0.0015	0.0052	0.0006	0.0004	0.0017	0.001	0.001	0.0009	0.0101	0.0057	0.0007	0.0006	0.0016	0.0009	0.0015	0.0015	0.0026	0.0031	0.0004	0.0007	0.0009	0.0012	0.0004	0.0028
0	0.0017	0.0001	0.0027	0	0.0001	0.0034	0.0006	0.0001	0	0.0063	0.0043	0.0005	0.0004	0.0016	0.0011	0.0005	0.0005	0.0007	0.0016	0.0002	0.0002	0.0005	0.0006	0.0003	0.0018
0.0002	0.0045	0.0009	0.0038	0.0008	0.0004	0.0007	0	0.0005	0.0002	0.0063	0.0042	0.0005	0.0004	0.001	0.0012	0.001	0.0005	0.0015	0.0015	0.0006	0.0007	0.0007	0.0005	0.0004	0.0019
0	0.0034	0.0002	0.0021	0.0002	0.0001	0.0003	0.0006	0	0	0.0075	0.0045	0.0008	0.0002	0.001	0.0004	0.001	0.0003	0.0006	0.0018	0.0005	0.0002	0.0007	0.0002	0	0.0021
0	0.0057	0.0001	0.0031	0.0005	0.0003	0.0005	0.0005	0.0005	0.0002	0.0072	0.0042	0.0004	0.0003	0.0011	0.001	0.0012	0.0009	0.0015	0.0019	0.0009	0.0004	0.001	0.0007	0.0002	0.0015
0.0003	0.0028	0	0.0023	0	0	0.0025	0	0	0	0.0054	0.0053	0.0003	0	0.0023	0.0009	0.0013	0.0009	0.0011	0.002	0.0002	0	0.0011	0.0003	0.0001	0.0011
0.001	0.0446	0.0394	0.0098	0.0012	0.0014	0.0017	0.0013	0.001	0.0021	0.011	0.0047	0.0009	0.0017	0.0043	0.0048	0.0778	0.0067	0.084	0.0061	0.0014	0.0014	0.0023	0.0017	0.0008	0.003
0.0001	0.0056	0.0007	0.0029	0.0005	0.0008	0.0012	0.0007	0.0004	0.0007	0.0056	0.0056	0.0008	0.0006	0.003	0.0011	0.0014	0.0006	0.0015	0.0026	0.0009	0.0006	0.0009	0.0011	0.0007	0.0024
0.0003	0.0075	0.0007	0.0036	0.0003	0.0008	0.001	0.0004	0.0005	0.0009	0.0054	0.0048	0.0006	0.0004	0.0024	0.0012	0.0014	0.0008	0.0013	0.0022	0.0005	0.0005	0.0007	0.0006	0.0005	0.0017
0	0.004	0.0002	0.0019	0.0003	0.0003	0.0006	0.0004	0.0003	0.0001	0.0043	0.0045	0.0009	0.0001	0.0019	0.001	0.0011	0.0008	0.0019	0.0018	0.0007	0.0006	0.0014	0.0009	0.0002	0.0017
0	0.0041	0	0.002	0.0007	0.0003	0.0009	0.0004	0.0003	0.0003	0.0048	0.0052	0.0005	0.0003	0.0022	0.001	0.001	0.0006	0.0016	0.002	0.0007	0.0005	0.0017	0.001	0.0005	0.0023
0.0008	0.0026	0.0001	0.0011	0.0001	0	0	0	0.0001	0	0.0018	0.0023	0.0001	0.0001	0.0007	0	0	0.0003	0.0005	0.0012	0.0001	0.0002	0.0005	0.0003	0	0.0007
0.0008	0.0062	0.0015	0.007	0.0019	0.0014	0.0063	0.0027	0.0014	0.0019	0.0125	0.0185	0.0021	0.0018	0.012	0.0031	0.0041	0.0034	0.0058	0.0071	0.0021	0.0015	0.0036	0.0029	0.002	0.0068
0.0008	0.0066	0.0008	0.0045	0.0007	0.001	0.0015	0.0014	0.0011	0.0012	0.0106	0.0084	0.0009	0.001	0.0043	0.0015	0.0026	0.0013	0.0025	0.0039	0.0009	0.0011	0.0021	0.0014	0.0009	0.0032
0.0022	0.0048	0.0015	0.0048	0.001	0.0015	0.0023	0.0015	0.0008	0.0012	0.0098	0.0064	0.0008	0.0008	0.0024	0.0015	0.0033	0.0035	0.0045	0.0036	0.0019	0.0024	0.0033	0.0017	0.0009	0.0026
0.0006	0.0054	0.0008	0.0034	0.0009	0.0009	0.0018	0.001	0.0007	0.0008	0.0074	0.005	0.0009	0.0006	0.0027	0.0012	0.0021	0.0009	0.0022	0.0028	0.0008	0.0009	0.0019	0.0017	0.001	0.0024
0.0003	0.0075	0.0007	0.0032	0.0007	0.0007	0.001	0.0007	0.0007	0.0007	0.0088	0.0067	0.0008	0.0005	0.0032	0.0012	0.002	0.0009	0.0023	0.0036	0.0008	0.0007	0.0011	0.0009	0.0006	0.0029
0.0005	0.0078	0.0009	0.0045	0.0006	0.0013	0.0016	0.0012	0.001	0.0008	0.01	0.0076	0.0011	0.0008	0.0032	0.0012	0.0024	0.0013	0.0031	0.0043	0.0011	0.0009	0.002	0.0012	0.001	0.0035
0.0002	0.004	0.0005	0.002	0.0002	0.0004	0.0009	0.0007	0.0002	0.0005	0.0043	0.0041	0.0005	0.0006	0.0015	0.0006	0.001	0.0004	0.0009	0.0017	0.0007	0.0004	0.0004	0.0006	0.0004	0.0013
0.0003	0.0053	0.0005	0.0027	0.0003	0.0002	0.001	0.0008	0.0007	0.0004	0.0069	0.0056	0.0005	0.0004	0.0023	0.0008	0.0016	0.0007	0.0024	0.0028	0.0009	0.0004	0.0012	0.0013	0.0006	0.0021
0.0003	0.0069	0.0005	0.0034	0.0004	0.0003	0.0008	0.0005	0.0006	0.0009	0.006	0.0046	0.0004	0.0004	0.0015	0.0014	0.0014	0.0008	0.0011	0.0022	0.0004	0.0006	0.0008	0.0004	0.0004	0.0018
0.0011	0.0071	0.0028	0.0064	0.0006	0.0006	0.0008	0.0117	0.014	0.0031	0.0088	0.0054	0.0011	0.0012	0.0022	0.0022	0.006	0.0017	0.0245	0.0258	0.0025	0.0016	0.0032	0.0011	0.0005	0.0027
0.0001	0.0043	0.0006	0.0021	0.0003	0.0005	0.0002	0.0003	0.0002	0.0003	0.0044	0.0034	0.0005	0.0005	0.0009	0.0008	0.0008	0.0004	0.001	0.0016	0.0001	0.0004	0.0008	0.0002	0.0004	0.0013
0.0003	0.0067	0.0004	0.0032	0.0004	0.0004	0.0007	0.0005	0.0006	0.0006	0.0079	0.0043	0.0005	0.0001	0.0013	0.0012	0.0009	0.0007	0.0018	0.0017	0.0003	0	0.0009	0.0006	0.0003	0.0018
0.0004	0.007	0.0002	0.0033	0.0003	0.0006	0.0016	0.0006	0.0009	0.0006	0.0057	0.0045	0.0003	0.0007	0.0014	0.0012	0.0011	0.0004	0.0014	0.0021	0.0004	0.0001	0.0007	0.0004	0.0002	0.0015
0.0001	0.0059	0.0006	0.0036	0.0006	0.0007	0.0007	0.0006	0.0005	0.0006	0.0095	0.0048	0.0005	0.0003	0.001	0.0013	0.0007	0.0007	0.001	0.0025	0.0001	0.0002	0.0005	0.0011	0.0005	0.0021
0.0007	0.0071	0.0007	0.0057	0.001	0.0008	0.0061	0.0014	0.0008	0.001	0.0152	0.0082	0.001	0.0009	0.0032	0.0017	0.0022	0.0027	0.0026	0.0047	0.001	0.0005	0.0021	0.0018	0.0009	0.0043
0.0006	0.01	0.0009	0.0047	0.0009	0.0011	0.0018	0.0013	0.0011	0.0011	0.0129	0.008	0.0006	0.001	0.0037	0.0013	0.0023	0.001	0.0034	0.0044	0.0008	0.0006	0.0018	0.0015	0.001	0.0035
0.0007	0.0075	0.0007	0.0042	0.0006	0.0007	0.0013	0.0013	0.0008	0.0008	0.0098	0.006	0.0008	0.0008	0.0031	0.0012	0.0022	0.0013	0.0022	0.0039	0.0006	0.0007	0.0023	0.0015	0.0008	0.0031
0.0001	0.0063	0.0005	0.003	0.0004	0.0006	0.0005	0.0007	0.0003	0.0003	0.0074	0.0048	0.0005	0.0005	0.0016	0.0009	0.0013	0.0007	0.0019	0.0027	0.0006	0.0008	0.0012	0.0007	0.0006	0.0028
0	0.0033	0	0.0017	0.0002	0	0.0003	0	0.0001	0	0.004	0.003	0.0004	0	0.0008	0.0005	0.0008	0.0001	0.0014	0.0014	0.0004	0.0002	0.0008	0.0005	0.0001	0.0008
0.0007	0.0081	0.0011	0.0053	0.001	0.0008	0.0017	0.0016	0.0009	0.0011	0.0109	0.005	0.0005	0.0006	0.0023	0.0012	0.002	0.0008	0.002	0.003	0.0008	0.0007	0.0015	0.0014	0.0007	0.0029
0.0033	0.0123	0.0033	0.0076	0.0007	0.0008	0.0016	0.0007	0.001	0.0013	0.0162	0.0099	0.0009	0.0009	0.0032	0.0018	0.0064	0.0013	0.0031	0.0037	0.0009	0.0009	0.0052	0.0013	0.0007	0.003
0.0003	0.0052	0.0003	0.0028	0.0002	0.0008	0.001	0.0007	0.0003	0.0006	0.0064	0.0051	0.0008	0.0006	0.0021	0.001	0.0019	0.0009	0.0017	0.0033	0.0013	0.0007	0.0022	0.001	0.001	0.0026
0	0.0035	0.0004	0.0021	0.0003	0.0005	0.002	0.0006	0.0004	0.0004	0.0057	0.0065	0.0009	0.0005	0.0038	0.0014	0.0017	0.0011	0.0023	0.003	0.0006	0.0007	0.0013	0.0008	0.0006	0.003
0.0006	0.0108	0.001	0.0049	0.0009	0.001	0.0018	0.0014	0.0012	0.0012	0.0124	0.0085	0.0008	0.001	0.0038	0.0016	0.0027	0.0014	0.0028	0.0041	0.0011	0.0013	0.0029	0.0019	0.0013	0.004
0.0005	0.0069	0.0012	0.0033	0.0008	0.0008	0.0014	0.0008	0.0004	0.0009	0.0063	0.0043	0.0006	0.0006	0.0015	0.0012	0.0021	0.0008	0.0026	0.0021	0.0007	0.0005	0.0014	0.0009	0.0004	0.0017
0.0004	0.0089	0.001	0.0042	0.0011	0.0008	0.0011	0.0013	0.0009	0.0009	0.0095	0.0075	0.0011	0.0007	0.0029	0.0019	0.0018	0.0013	0.0026	0.0034	0.0008	0.0005	0.0018	0.0018	0.0008	0.003

Table 54. NPC LDD Dataset encompassing the 65 rows and 143 features peaks values.

3.4. Plot of NPC Liver Dysfunction Disease Dataset

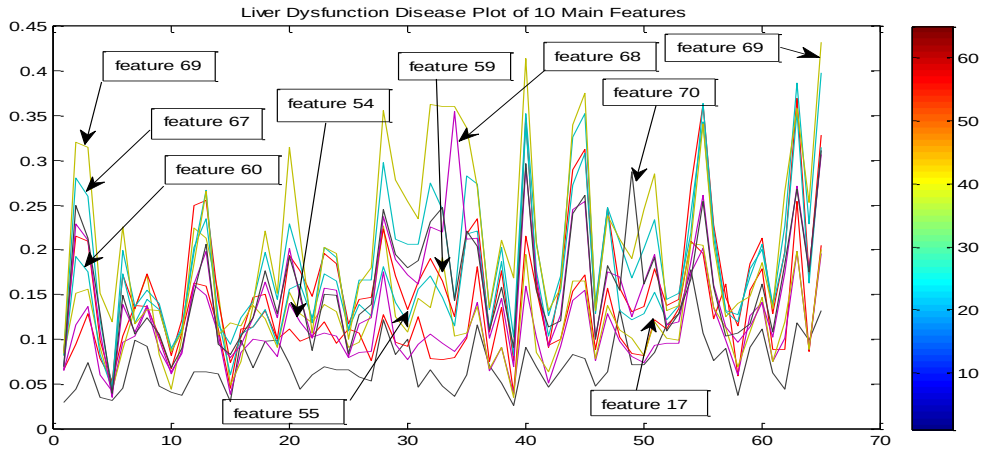

Figure 30. Plot of the 10 most important features (69, 68, 70, 67, 58, 59, 60, 55, 54, 17) for NPC disease-associated liver dysfunction. They correspond to the following biomolecules glutamate, glutamine, taurine, lactate, glycerophosphocholine, myo-Inositol, alanine, leucine, isoleucine, etc., considered as potential biomarkers in the NPC LDD diagnosis.

3.5. Contour Plot of NPC Liver Dysfunction Disease (NPC LDD) Dataset

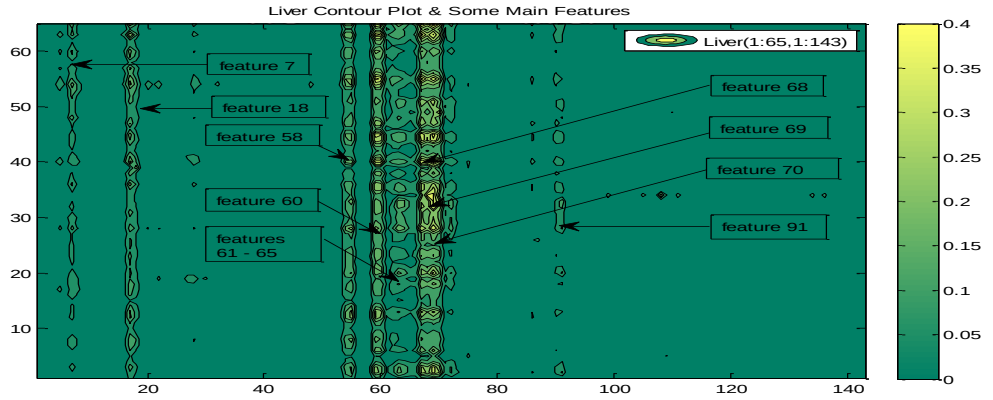

Figure 31. Contour plot showing most important features (69, 68, 70, 67, 58, 59, 60, 55, 54, 17, 18, 7) for NPC disease- associated liver dysfunction diagnosis. They correspond to the following biomolecules glutamate, glutamine, taurine, lactate, glycerophosphocholine, myo-Inositol, alanine, leucine, isoleucine, etc., considered as potential biomarkers in the liver dysfunction and associated disease diagnosis.

3.6. Scale Data Plot of NPC Liver Dysfunction Disease (NPC LDD) Dataset

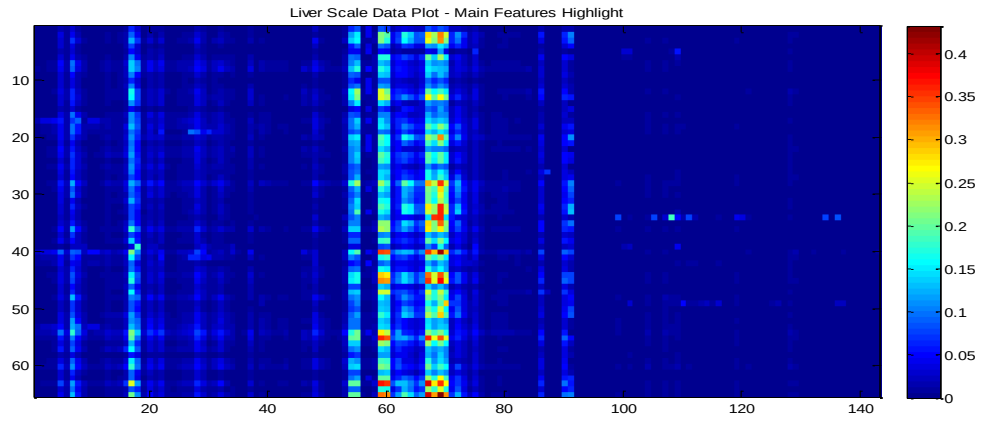

Figure 32. Scale data plot showing significant changes of intensity for the main features 69, 68, 70, 67, 58, 59, 60, 55, 54, 17, 18, 7 – They correspond to the following potential biomolecules glutamate, glutamine, taurine, lactate, glycerophosphocholine, myo-Inositol, alanine, leucine, isoleucine, etc., in the NPC disease-associated liver dysfunction.

3.7. Scale Data Plot of NPC Liver Dysfunction Disease (NPC LDD) Dataset

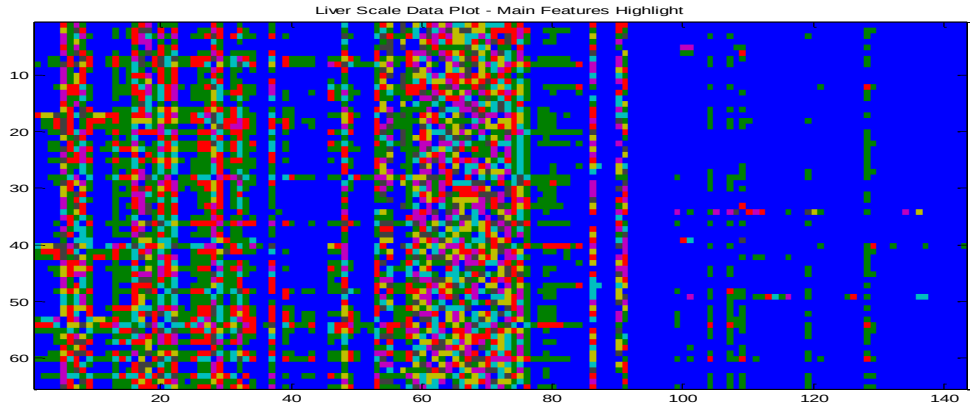

Figure 33. Scale data image showing significant changes of intensity for the major features 69, 68, 70, 67, 58, 59, 60, 55, 54, 17, 18, 7 – They correspond to the following biomolecules glutamate, glutamine, taurine, lactate, glycerophosphocholine, myo-Inositol, alanine, leucine, isoleucine, etc., considered as potential biomarkers in the liver dysfunction and associated disease diagnosis.

3.8. Contour Plot of NPC Liver Dysfunction Disease (NPC LDD) Dataset

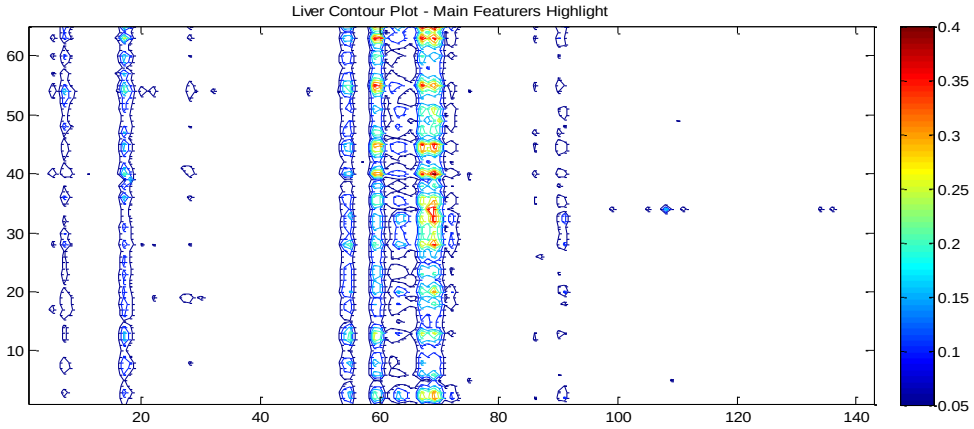

Figure 34. Contour plot showing significant changes of intensity for major features 69, 68, 70, 67, 58, 59, 60, 55, 54, 17, 18, 7 –They correspond to the following biomolecules glutamate, glutamine, taurine, lactate, glycerophosphocholine, myo-Inositol, alanine, leucine, isoleucine, etc., considered as potential biomarkers in the liver dysfunction and associated disease diagnosis.

3.9. Boxplot of NPC Liver Dysfunction Disease (NPC LDD) Dataset

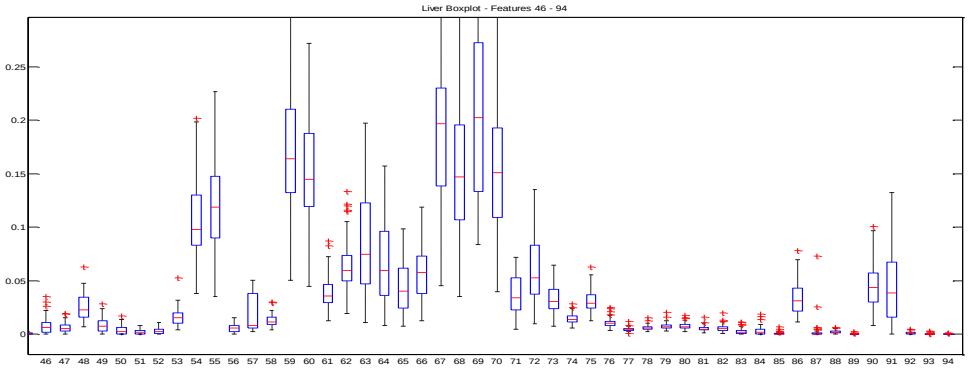

Figure 35. Zoomed boxplot showing features numbers 48-76 and presenting the features 69, 67, 59, 70, 68, 60, 55, 54, 63, 64, 66, 65, 90, 91 as the most important in the NPC-associated liver dysfunction. They correspond to the following potential biomarkers glutamate, glutamine, taurine, lactate, glycerophosphocholine, myo-Inositol, gamma-phosphorylcholine, leucine, isoleucine, alanine, etc., considered as most important ones in the NPC LDD diagnosis according to the boxplot technique.

3.10. Filled Contour Plot of NPC LDD Dataset

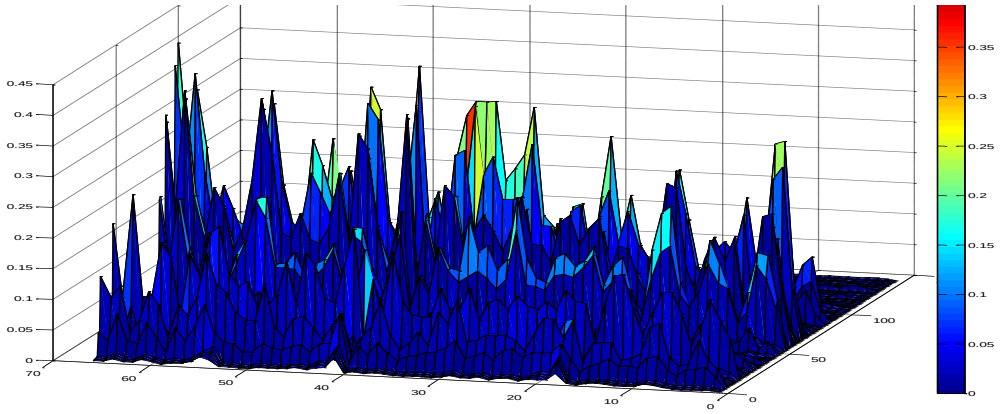

Figure 36. Filled contour plot showing a 3D representation of the NPC-associated liver dysfunction dataset with more intense peaks representing the most important ones.